```
Db        594 EIPQFIGRSYLTYDNPDILKRVSG---SRS------NVFMRFKTTAKDGLLLWRG--DSP 642

Qy        146 AGQGKDFISLGLQDGHLVFRYQLGSGEARLVSEDPINDGEWHRVTALREGRRGSIQVDGE 205
                   |||||||:|| ||| | |||| | ::     ||| |||| |:|:|: | | ||
Db        643 MRPNSDFISLGLRDGALVFSYNLGSGVASIMVNGSFNDGRWHRVKAVRDGQSGKITVDDY 702

Qy        206 ELVSGRSPGPNVAVNAKGSVYIGGAPDVATLTGGRFSSGITGCVKNLVLHSARPGAPPPQ 265
                 :|:|||    :|  |::|:||  ::|  |  ::  |: ||: :  | :
Db        703 GARTGKSPGMMRQLNINGALYVGGMKEIALHTNRQYMRGLVGCISHFTLST-------DY 755

Qy        266 PLDLQHRAQAGANTRPC 282
               : |   |  | |   |
Db        756 HISLVEDAVDGKNINTC 772

RESULT 6
Q80WX4
ID   Q80WX4      PRELIMINARY;      PRT;   295 AA.
AC   Q80WX4;
DT   01-JUN-2003 (TrEMBLrel. 24, Created)
DT   01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Similar to Agrin (Fragment).
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Olfactory epithelium;
RX   MEDLINE=22388257; PubMed=12477932;
RA   Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA   Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA   Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA   Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA   Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA   Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA   Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA   Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA   Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA   Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA   Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA   Fahey J., Helton E., Ketteman M., Madan A., Rodrigues S., Sanchez A.,
RA   Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA   Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA   Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA   Krzywinski M.I., Skalska U., Smailus D.E., Schnerch A., Schein J.E.,
RA   Jones S.J., Marra M.A.;
RT   "Generation and initial analysis of more than 15,000 full-length human
RT   and mouse cDNA sequences.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Olfactory epithelium;
RA   Strausberg R.;
RL   Submitted (APR-2003) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; BC051455; AAH51455.1; -.
DR   GO; GO:0005509; F:calcium ion binding; IEA.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR000742; EGF_2.
DR   InterPro; IPR001881; EGF_Ca.
DR   InterPro; IPR006209; EGF_like.
DR   InterPro; IPR006210; IEGF.
DR   InterPro; IPR001791; Laminin_G.
DR   Pfam; PF00008; EGF; 1.
DR   Pfam; PF00054; laminin_G; 1.
DR   SMART; SM00181; EGF; 1.
DR   SMART; SM00179; EGF_CA; 1.
DR   SMART; SM00282; LamG; 1.
DR   PROSITE; PS00022; EGF_1; 1.
DR   PROSITE; PS01186; EGF_2; 1.
DR   PROSITE; PS50025; LAM_G_DOMAIN; 2.
FT   NON_TER       1      1
```

```
SQ   SEQUENCE   295 AA;  32046 MW;  6165DC666EEC885C CRC64;

  Query Match             20.8%;  Score 325;  DB 11;  Length 295;
  Best Local Similarity   30.8%;  Pred. No. 1.4e-17;
  Matches   78;  Conservative   38;  Mismatches  105;  Indels   32;  Gaps    6;

Qy         34 EHEENPCQLREPCLHGGTC----QGTRCLCLPGFSGPRCQQGSGHGIAESDWHLEGSGGN 89
              |:  :|| :  || |||:|    :|  | |  || |  ||:     |
Db         68 ENAAHPC-VGAPCAHGGSCRPRKEGYECDCPLGFEGLNCQKAIIEAI------------- 113

Qy         90 DAPGQYGAYFHDDGFLAFPGHVFSRSLPEVPETIELEVRTSTASGLLLWQGVEVGEAGQG 149
              : |      |    :| :      : :        :  :|:   |||||:|
Db        114 EIP-----QFIGRSYLTYDNPNILKRVSGSRSNAFMRFKTTAKDGLLLWRG--DSPMRPN 166

Qy        150 KDFISLGLQDGHLVFRYQLGSGEARLVSEDPINDGEWHRVTALREGRRGSIQVDGEELVS 209
               |||||||:|| |:| | |||| | ::     :|| |||| |:|:|: | | ||     :
Db        167 SDFISLGLRDGALIFSYNLGSGVASIMVNGSFSDGRWHRVKAVRDGQSGKITVDDYGART 226

Qy        210 GRSPGPNVAVNAKGSVYIGGAPDVATLTGGRFSSGITGCVKNLVLHSARPGAPPPQPLDL 269
              |:|||    :|  |::|:||  ::|  |  ::  |: ||: :  | :           :|
Db        227 GKSPGLMRQLNINGALYVGGMKEIALHTNRQYLRGLVGCISHFTLST-------DYHISL 279

Qy        270 QHRAQAGANTRPC 282
                 |  | |   |
Db        280 VEDAVDGKNINTC 292

RESULT 7
Q8BGP3
ID   Q8BGP3      PRELIMINARY;      PRT;  1009 AA.
AC   Q8BGP3;
DT   01-MAR-2003 (TrEMBLrel. 23, Created)
DT   01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Hypothetical EGF-like domain.
GN   AU040377.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6J; TISSUE=Forelimb, Skin, Testis, and Thymus;
RX   MEDLINE=22354683; PubMed=12466851;
RA   The FANTOM Consortium,
RA   the RIKEN Genome Exploration Research Group Phase I & II Team;
RT   "Analysis of the mouse transcriptome based on functional annotation of
RT   60,770 full-length cDNAs.";
RL   Nature 420:563-573(2002).
DR   EMBL; AK031136; BAC27271.1; -.
DR   EMBL; AK033332; BAC28235.1; -.
DR   EMBL; AK037223; BAC29762.1; -.
DR   EMBL; AK041546; BAC30982.1; -.
DR   MGD; MGI:2146149; AU040377.
DR   GO; GO:0005509; F:calcium ion binding; IEA.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR000742; EGF_2.
DR   InterPro; IPR001881; EGF_Ca.
DR   InterPro; IPR006209; EGF_like.
DR   InterPro; IPR003962; FnIII_subd.
DR   InterPro; IPR003961; FN_III.
DR   InterPro; IPR008957; FN_III-like.
DR   InterPro; IPR006210; IEGF.
DR   InterPro; IPR001791; Laminin_G.
DR   Pfam; PF00008; EGF; 2.
DR   Pfam; PF00041; fn3; 2.
DR   Pfam; PF00054; laminin_G; 3.
DR   PRINTS; PR00014; FNTYPEIII.
DR   SMART; SM00181; EGF; 3.
DR   SMART; SM00179; EGF_CA; 2.
DR   SMART; SM00060; FN3; 2.
DR   SMART; SM00282; LamG; 3.
DR   PROSITE; PS00022; EGF_1; 3.
```

```
DR   PROSITE; PS01186; EGF_2; 2.
DR   PROSITE; PS50025; LAM_G_DOMAIN; 3.
KW   Hypothetical protein.
SQ   SEQUENCE   1009 AA;  109834 MW;  80DD954F255C102B CRC64;

  Query Match             20.8%;  Score 325;  DB 11;  Length 1009;
  Best Local Similarity   30.8%;  Pred. No. 6.9e-17;
  Matches   78;  Conservative   38;  Mismatches  105;  Indels   32;  Gaps    6;

Qy         34 EHEENPCQLREPCLHGGTC----QGTRCLCLPGFSGPRCQQGSGHGIAESDWHLEGSGGN 89
              |:  :|| :  || |||:|    :|  | |  || |  ||:     |
Db        782 ENAAHPC-VGAPCAHGGSCRPRKEGYECDCPLGFEGLNCQKAIIEAI------------- 827

Qy         90 DAPGQYGAYFHDDGFLAFPGHVFSRSLPEVPETIELEVRTSTASGLLLWQGVEVGEAGQG 149
              : |      |    :| :      : :        :  :|:   |||||:|
Db        828 EIP-----QFIGRSYLTYDNPNILKRVSGSRSNAFMRFKTTAKDGLLLWRG--DSPMRPN 880

Qy        150 KDFISLGLQDGHLVFRYQLGSGEARLVSEDPINDGEWHRVTALREGRRGSIQVDGEELVS 209
               |||||||:|| |:| | |||| | ::     :|| |||| |:|:|: | | ||     :
Db        881 SDFISLGLRDGALIFSYNLGSGVASIMVNGSFSDGRWHRVKAVRDGQSGKITVDDYGART 940

Qy        210 GRSPGPNVAVNAKGSVYIGGAPDVATLTGGRFSSGITGCVKNLVLHSARPGAPPPQPLDL 269
              |:|||    :|  |::|:||  ::|  |  ::  |: ||: :  | :          : |
Db        941 GKSPGLMRQLNINGALYVGGMKEIALHTNRQYLRGLVGCISHFTLST-------DYHISL 993

Qy        270 QHRAQAGANTRPC 282
                 |  | |   |
Db        994 VEDAVDGKNINTC 1006

RESULT 8
Q80V56
ID   Q80V56      PRELIMINARY;      PRT;   263 AA.
AC   Q80V56;
DT   01-JUN-2003 (TrEMBLrel. 24, Created)
DT   01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Agrn protein (Fragment).
GN   AGRN.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=FVB/N; TISSUE=Breast tumor;
RX   MEDLINE=22388257; PubMed=12477932;
RA   Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA   Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA   Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA   Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA   Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA   Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA   Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA   Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA   Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA   Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA   Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA   Fahey J., Helton E., Ketteman M., Madan A., Rodrigues S., Sanchez A.,
RA   Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA   Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA   Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA   Krzywinski M.I., Skalska U., Smailus D.E., Schnerch A., Schein J.E.,
RA   Jones S.J., Marra M.A.;
RT   "Generation and initial analysis of more than 15,000 full-length human
RT   and mouse cDNA sequences.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=FVB/N; TISSUE=Breast tumor;
RA   Strausberg R.;
RL   Submitted (JAN-2003) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; BC043318; AAH43318.1; -.
DR   GO; GO:0005509; F:calcium ion binding; IEA.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR001881; EGF_Ca.
DR   InterPro; IPR006209; EGF_like.
DR   InterPro; IPR006210; IEGF.
DR   InterPro; IPR001791; Laminin_G.
DR   Pfam; PF00008; EGF; 1.
DR   Pfam; PF00054; laminin_G; 1.
DR   SMART; SM00181; EGF; 1.
DR   SMART; SM00179; EGF_CA; 1.
DR   SMART; SM00282; LamG; 1.
DR   PROSITE; PS00022; EGF_1; 1.
DR   PROSITE; PS01186; EGF_2; 1.
DR   PROSITE; PS50025; LAM_G_DOMAIN; 1.
FT   NON_TER       1      1
SQ   SEQUENCE   263 AA;  28395 MW;  89708020EBFBD515 CRC64;

  Query Match             19.4%;  Score 304;  DB 11;  Length 263;
  Best Local Similarity   31.1%;  Pred. No. 5.6e-16;
  Matches   84;  Conservative   38;  Mismatches   98;  Indels   50;  Gaps    9;

Qy         31 DLCEHEENPC--QLREPCLHGGTC---QGT-RCLCLPGFSGPRCQQGSGHGIAESDWHLE 84
              |:     :||   :  |||:||:|   : |  |||  ||||  |::|           :|
Db         27 DVAPFAGHPCTQAVDNPCLNGGSCIPREATYECLCPGGFSGLHCEKGI----------VE 76

Qy         85 GSGGNDAPGQYGAYFHDDGFLAFPGHVFSRSLPEVPET--------IELEVRTSTASGLL 136
               | |            |   ||| |  :   |  | |:         || :||    ||:
Db         77 KSVG------------DLETLAFDGRTYIEYLNAVTESEKALQSNHFELSLRTEATQGLV 124

Qy        137 LWQGVEVGEAGQGKDFISLGLQDGHLVFRYQLGSGEARLVSEDPINDGEWHRVTALREGR 196
              ||    :|: |:  |:::| : ||||   | |||    | |   :|   | || | || |
Db        125 LW----IGKVGERADYMALAIVDGHLQLSYDLGSQPVVLRSTVKVNTNRWLRVRAHREHR 180

Qy        197 RGSIQVDGEELVSGRSPGPNVAVNAKGSVYIGGAP--DVATLTGGRFSSGITGCVKNLVL 254
               ||:||  |  |:| ||     ::  |::::||     |       : :|  ||::::|:
Db        181 EGSLQVGNEAPVTGSSPLGATQLDTDGALWLGGLQKLPVGQALPKAYGTGFVGCLRDVVV 240

Qy        255 HSARPGAPPPQPLDLQHRAQAGANTRPCPS 284
                        : | |   |      ||||:
Db        241 GH--------RQLHLLEDAVTKPELRPCPT 262

RESULT 9
Q99KT4
ID   Q99KT4      PRELIMINARY;      PRT;   263 AA.
AC   Q99KT4;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Hypothetical protein (Fragment).
GN   AGRN.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Strausberg R.;
RL   Submitted (FEB-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; BC004020; AAH04020.1; -.
DR   HSSP; P00740; 1EDM.
DR   MGD; MGI:87961; Agrn.
DR   GO; GO:0007009; P:plasma membrane organization and biogenesis; IMP.
DR   GO; GO:0007268; P:synaptic transmission; IDA.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR006209; EGF_like.
DR   InterPro; IPR006210; IEGF.
DR   InterPro; IPR001791; Laminin_G.
DR   Pfam; PF00008; EGF; 1.
DR   Pfam; PF00054; laminin_G; 1.
DR   SMART; SM00181; EGF; 1.
DR   SMART; SM00282; LamG; 1.
DR   PROSITE; PS00022; EGF_1; 1.
```

```
DR   PROSITE; PS01186; EGF_2; 1.
DR   PROSITE; PS50025; LAM_G_DOMAIN; 1.
KW   Hypothetical protein; EGF-like domain.
FT   NON_TER       1      1
SQ   SEQUENCE   263 AA;  28381 MW;  F9739020EBFA9040 CRC64;

  Query Match             19.4%;  Score 304;  DB 11;  Length 263;
  Best Local Similarity   31.1%;  Pred. No. 5.6e-16;
  Matches   84;  Conservative   38;  Mismatches   98;  Indels   50;  Gaps    9;

Qy          31 DLCEHEENPC--QLREPCLHGGTC---QGT-RCLCLPGFSGPRCQQGSGHGIAESDWHLE 84
               |:     :||   :  |||:||:|   : |  |||  ||||  |::|           :|
Db          27 DVAPFAGHPCTQAVDNPCLNGGSCIPREATYECLCPGGFSGLHCEKGI----------VE 76

Qy          85 GSGGNDAPGQYGAYFHDDGFLAFPGHVFSRSLPEVPET--------IELEVRTSTASGLL 136
                | |            |   ||| |  :   |  | |:         || :||    ||:
Db          77 KSVG------------DLETLAFDGRTYIEYLNAVTESEKALQSNHFELSLRTEATQGLV 124

Qy         137 LWQGVEVGEAGQGKDFISLGLQDGHLVFRYQLGSGEARLVSEDPINDGEWHRVTALREGR 196
               ||    :|: |:  |:::| : ||||   | |||    | |   :|   | || | || |
Db         125 LW----IGKVGERADYMALAIVDGHLQLSYDLGSQPVVLRSTVKVNTNRWLRVRAHREHR 180

Qy         197 RGSIQVDGEELVSGRSPGPNVAVNAKGSVYIGGAP--DVATLTGGRFSSGITGCVKNLVL 254
                ||:||  |  |:| ||     ::  |::::||     |       : :|  ||::::|:
Db         181 EGSLQVGNEAPVTGSSPLGATQLDTDGALWLGGLQKLPVGQALPKAYGTGFVGCLRDVVV 240

Qy         255 HSARPGAPPPQPLDLQHRAQAGANTRPCPS 284
                         : | |   |      ||||:
Db         241 GH--------RQLHLLEDAVTKPELRPCPT 262


RESULT 10
Q8K326
ID   Q8K326      PRELIMINARY;      PRT;   350 AA.
AC   Q8K326;
DT   01-OCT-2002 (TrEMBLrel. 22, Created)
DT   01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Hypothetical protein (Fragment).
GN   AGRN.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Strausberg R.;
RL   Submitted (MAY-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; BC028992; AAH28992.1; -.
DR   MGD; MGI:87961; Agrn.
DR   GO; GO:0007009; P:plasma membrane organization and biogenesis; IMP.
DR   GO; GO:0007268; P:synaptic transmission; IDA.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR001881; EGF_Ca.
DR   InterPro; IPR006209; EGF_like.
DR   InterPro; IPR006210; IEGF.
DR   InterPro; IPR001791; Laminin_G.
DR   Pfam; PF00008; EGF; 1.
DR   Pfam; PF00054; laminin_G; 2.
DR   SMART; SM00181; EGF; 1.
DR   SMART; SM00179; EGF_CA; 1.
DR   SMART; SM00282; LamG; 2.
DR   PROSITE; PS00022; EGF_1; 1.
DR   PROSITE; PS01186; EGF_2; 1.
DR   PROSITE; PS50025; LAM_G_DOMAIN; 2.
KW   Hypothetical protein; EGF-like domain.
FT   NON_TER       1      1
SQ   SEQUENCE   350 AA;  37809 MW;  020ED7F15A353ED0 CRC64;

  Query Match             19.4%;  Score 304;  DB 11;  Length 350;
  Best Local Similarity   31.1%;  Pred. No. 8e-16;
  Matches   84;  Conservative   38;  Mismatches   98;  Indels   50;  Gaps    9;

Qy          31 DLCEHEENPC--QLREPCLHGGTC---QGT-RCLCLPGFSGPRCQQGSGHGIAESDWHLE 84
               |:     :||   :  |||:||:|   : |  |||  ||||  |::|           :|
Db         114 DVAPFAGHPCTQAVDNPCLNGGSCIPREATYECLCPGGFSGLHCEKGI----------VE 163

Qy          85 GSGGNDAPGQYGAYFHDDGFLAFPGHVFSRSLPEVPET--------IELEVRTSTASGLL 136
                | |            |   ||| |  :   |  | |:         || :||    ||:
Db         164 KSVG------------DLETLAFDGRTYIEYLNAVTESEKALQSNHFELSLRTEATQGLV 211

Qy         137 LWQGVEVGEAGQGKDFISLGLQDGHLVFRYQLGSGEARLVSEDPINDGEWHRVTALREGR 196
               ||    :|: |:  |:::| : ||||   | |||    | |   :|   | || | || |
Db         212 LW----IGKVGERADYMALAIVDGHLQLSYDLGSQPVVLRSTVKVNTNRWLRVRAHREHR 267

Qy         197 RGSIQVDGEELVSGRSPGPNVAVNAKGSVYIGGAP--DVATLTGGRFSSGITGCVKNLVL 254
                ||:||  |  |:| ||     ::  |::::||     |       : :|  ||::::|:
Db         268 EGSLQVGNEAPVTGSSPLGATQLDTDGALWLGGLQKLPVGQALPKAYGTGFVGCLRDVVV 327

Qy         255 HSARPGAPPPQPLDLQHRAQAGANTRPCPS 284
                         : | |   |      ||||:
Db         328 GH--------RQLHLLEDAVTKPELRPCPT 349


RESULT 11
Q86SD6
ID   Q86SD6      PRELIMINARY;      PRT;   478 AA.
AC   Q86SD6;
DT   01-JUN-2003 (TrEMBLrel. 24, Created)
DT   01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Perlecan homologue (Fragment).
GN   CI-PERLECAN.
OS   Ciona intestinalis.
OC   Eukaryota; Metazoa; Chordata; Urochordata; Ascidiacea; Enterogona;
OC   Phlebobranchia; Cionidae; Ciona.
OX   NCBI_TaxID=7719;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Hotta K., Takahashi H., Satoh N.;
RT   "Further Characterization of Brachyury-Downstream Genes in Ciona
RT   intestinalis Embryo.";
RL   Submitted (DEC-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AB076890; BAC57513.1; -.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR006209; EGF_like.
DR   InterPro; IPR001791; Laminin_G.
DR   Pfam; PF00054; laminin_G; 2.
DR   SMART; SM00282; LamG; 2.
DR   PROSITE; PS00022; EGF_1; 1.
DR   PROSITE; PS01186; EGF_2; 1.
DR   PROSITE; PS50025; LAM_G_DOMAIN; 2.
FT   NON_TER       1      1
SQ   SEQUENCE   478 AA;  52039 MW;  5D55465CC835B82E CRC64;

  Query Match             18.8%;  Score 295;  DB 5;  Length 478;
  Best Local Similarity   28.7%;  Pred. No. 6e-15;
  Matches   80;  Conservative   34;  Mismatches   77;  Indels   88;  Gaps    9;

Qy          25 RDGFKGDL-----------------------CEHEENPCQLREPCLHGGTCQGTRCLCLP 61
               |||| | :                       || :  |    |  :             |
Db         237 RDGFDGCIDELRIDDIEIPFNEMFVRKNTRSCEVQPQPSVQDEAAV------------FP 284

Qy          62 GFSGPRCQQGSGHGIAESDWHLEGSGGNDAPGQYGAYFHDDGFLAFPGHVFSRSLPEVPE 121
               | | ||              ||             | |  | | :|    |        |
Db         285 GVSPPR--------------HL-------------ARFSGDSFASFEPTTFPHD--GSAE 315

Qy         122 TIELEVRTSTASGLLLWQGVEV-------------GEAGQGKDFISLGLQDGHLVFRYQL 168
                |  ::| ::  ||:|: | ::                |: ||:||:|:|:  ||: : |
Db         316 QISFDMRATSPDGLVLFHGEDLTSEPSETSRLHKRSHRGRSKDYISVGIQESKLVYSFDL 375

Qy         169 GSGEARLVSEDPINDGEWHRVTALREGRRGSIQV----DGEELVSGRSPGPNVAVNAKGS 224
               | |: | ||: |: || || || :   ||||: :    |    |: ::|| :   |  |
Db         376 GGGQGRAVSDFPVVDGLWHHVTIV---RRGSLALMWLDDVPTPVAAQAPGSHTMANTPGR 432

Qy         225 VYIGGAPDVATLTGGRFSSGITGCVKNLVLHSARPGAPP 263
```

```
              ||:||||||   | ||:::|  ||: ||       | ||
Db        433 VYVGGAPDVIFATAGRYTTGFMGCIGNLRF----TGIPP 467

RESULT 12
Q9V7I4
ID   Q9V7I4      PRELIMINARY;      PRT;  1361 AA.
AC   Q9V7I4;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   CG8403 protein.
GN   SP2353 OR CG8403.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC   Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Berkeley;
RX   MEDLINE=20196006; PubMed=10731132;
RA   Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA   Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA   George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA   Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA   Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA   Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA   Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA   Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA   Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA   Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA   Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA   Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA   de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA   Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA   Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA   Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA   Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA   Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA   Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA   Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA   Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA   Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA   Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA   Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA   Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA   Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA   Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA   Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA   Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA   Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA   Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA   Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA   Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA   Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA   Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA   Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT   "The genome sequence of Drosophila melanogaster.";
RL   Science 287:2185-2195(2000).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Celniker S.E., Adams M.D., Kronmiller B., Wan K.H., Holt R.A.,
RA   Evans C.A., Gocayne J.D., Amanatides P.G., Brandon R.C., Rogers Y.,
RA   Banzon J., An H., Baldwin D., Banzon J., Beeson K.Y., Busam D.A.,
RA   Carlson J.W., Center A., Champe M., Davenport L.B., Dietz S.M.,
RA   Dodson K., Dorsett V., Doup L.E., Doyle C., Dresnek D., Farfan D.,
RA   Ferriera S., Frise E., Galle R.F., Garg N.S., George R.A.,
RA   Gonzalez M., Houck J., Hoskins R.A., Hostin D., Howland T.J.,
RA   Ibegwam C., Jalali M., Kruse D., Li P., Mattei B., Moshrefi A.,
RA   McIntosh T.C., Moy M., Murphy B., Nelson C., Nelson K.A., Nunoo J.,
RA   Pacleb J., Paragas V., Park S., Patel S., Pfeiffer B.,
RA   Phouanenavong S., Pittman G.S., Puri V., Richards S., Scheeler F.,
```

```
RA   Stapleton M., Strong R., Svirskas R., Tector C., Tyler D.,
RA   Williams S.M., Zaveri J.S., Smith H.O., Venter J.C., Rubin G.M.;
RT   "Sequencing of Drosophila melanogaster genome.";
RL   Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RA   Misra S., Crosby M.A., Matthews B.B., Bayraktaroglu L., Campbell K.,
RA   Hradecky P., Huang Y., Kaminker J.S., Prochnik S.E., Smith C.D.,
RA   Tupy J.L., Bergman C., Berman B., Carlson J.W., Celniker S.E.,
RA   Clamp M., Drysdale R., Emmert D., Frise E., de Grey A., Harris N.,
RA   Kronmiller B., Marshall B., Millburn G., Richter J., Russo S.,
RA   Searle S.M.J., Smith E., Shu S., Smutniak F., Whitfield E.,
RA   Ashburner M., Gelbart W.M., Rubin G.M., Mungall C.J., Lewis S.E.;
RT   "Annotation of Drosophila melanogaster genome.";
RL   Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
RN   [4]
RP   SEQUENCE FROM N.A.
RA   Adams M.D., Celniker S.E., Gibbs R.A., Rubin G.M., Venter C.J.;
RL   Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
RN   [5]
RP   SEQUENCE FROM N.A.
RA   FlyBase;
RL   Submitted (SEP-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AE003808; AAF58071.2; -.
DR   HSSP; P08709; 1BF9.
DR   FlyBase; FBgn0034070; SP2353.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR006209; EGF_like.
DR   InterPro; IPR006210; IEGF.
DR   InterPro; IPR001791; Laminin_G.
DR   Pfam; PF00008; EGF; 1.
DR   Pfam; PF00054; laminin_G; 3.
DR   SMART; SM00181; EGF; 4.
DR   SMART; SM00282; LamG; 3.
DR   PROSITE; PS00022; EGF_1; 2.
DR   PROSITE; PS01186; EGF_2; 2.
DR   PROSITE; PS50025; LAM_G_DOMAIN; 3.
KW   EGF-like domain.
SQ   SEQUENCE   1361 AA;  150298 MW;  4C71EDA81A12ABED CRC64;

  Query Match             18.6%;  Score 292;  DB 5;  Length 1361;
  Best Local Similarity   35.0%;  Pred. No. 3.9e-14;
  Matches   64;  Conservative   37;  Mismatches   70;  Indels   12;  Gaps    5;

Qy         97 AYFH--DDGFLAFPGHVFSRSLPEVPETIELEVRTSTASGLLLWQGVEVGEAGQGKDFIS 154
              :|||  |   ::    | | |:      : | ::| : :|::|| | : |   :  |::|
Db       1186 SYFHYNDADTMS---QVISYSI-----DLNLRIKTHSENGVILWTGRQ-GTTEEHDDYLS 1236

Qy        155 LGLQDGHLVFRYQLGSGEARL-VSEDPINDGEWHRVTALREGRRGSIQVDGEELVSGRSP 213
              ||:: |:|.||| |||||  :  :   ::|| |||| |:|  : | ::||| : |: |:|
Db       1237 LGIEQGYLHFRYDLGSGEVDIRFNGTKVSDGLWHRVRAIRNSQEGYLEVDGRKTVTLRAP 1296

Qy        214 GPNVAVNAKGSVYIGGAPDVATLTGGRFSSGITGCVKNLVLHSARPGAPPPQPLDLQHRA 273
              |    :|    :|:|| |||   |  |: ||| ||:  :||         |  |  :|
Db       1297 GKLRQLNTDTGLYVGGMPDVGYFTHQRYFSGIVGCISEIVLAGEMKLNFDPNTLGTEHNV 1356

Qy        274 QAG 276
              : |
Db       1357 ETG 1359

RESULT 13
Q9NGV2
ID   Q9NGV2      PRELIMINARY;      PRT;  1361 AA.
AC   Q9NGV2;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   SP2353.
GN   SP2353 OR CG8403.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC   Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
```

```
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Serano T.L., Pendleton J.D., Rubin G.M.;
RT   "A reverse genetic screen for genes involved in Drosophila
RT   development.";
RL   Submitted (FEB-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF239610; AAF63502.1; -.
DR   HSSP; P08709; 1BF9.
DR   FlyBase; FBgn0034070; SP2353.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR006209; EGF_like.
DR   InterPro; IPR006210; IEGF.
DR   InterPro; IPR001791; Laminin_G.
DR   Pfam; PF00008; EGF; 1.
DR   Pfam; PF00054; laminin_G; 3.
DR   SMART; SM00181; EGF; 3.
DR   SMART; SM00282; LamG; 3.
DR   PROSITE; PS00022; EGF_1; 2.
DR   PROSITE; PS01186; EGF_2; 2.
DR   PROSITE; PS50025; LAM_G_DOMAIN; 3.
KW   EGF-like domain.
SQ   SEQUENCE   1361 AA;  150355 MW;  206D9F9BF9EDFE47 CRC64;

  Query Match             18.6%;  Score 292;  DB 5;  Length 1361;
  Best Local Similarity   35.0%;  Pred. No. 3.9e-14;
  Matches   64;  Conservative   37;  Mismatches   70;  Indels   12;  Gaps    5;

Qy          97 AYFH--DDGFLAFPGHVFSRSLPEVPETIELEVRTSTASGLLLWQGVEVGEAGQGKDFIS 154
               :|||  |   ::    | | |:      : | ::| : :|::|| | : |   :  |::|
Db        1186 SYFHYNDADTMS---QVISYSI-----DLNLRIKTHSENGVILWTGRQ-GTTEEHDDYLS 1236

Qy         155 LGLQDGHLVFRYQLGSGEARL-VSEDPINDGEWHRVTALREGRRGSIQVDGEELVSGRSP 213
               ||:: |:| ||| |||||  :  :   ::|| |||| |:|  : | ::||| : |: |:|
Db        1237 LGIEQGYLHFRYDLGSGEVDIRFNGTKVSDGLWHRVRAIRNSQEGYLEVDGRKTVTLRAP 1296

Qy         214 GPNVAVNAKGSVYIGGAPDVATLTGGRFSSGITGCVKNLVLHSARPGAPPPQPLDLQHRA 273
               |    :|    :|:|| |||   |  |: ||| ||:  :||         |  |  :|
Db        1297 GKLRQLNTDTGLYVGGMPDVGYFTHQRYFSGIVGCISEIVLAGEMKLNFDPNTLGTEHNV 1356

Qy         274 QAG 276
               : |
Db        1357 ETG 1359

RESULT 14
O75093
ID   O75093      PRELIMINARY;      PRT;  1534 AA.
AC   O75093;
DT   01-NOV-1998 (TrEMBLrel. 08, Created)
DT   01-AUG-1999 (TrEMBLrel. 11, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Slit-1 protein.
GN   SLIT-1.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=99033071; PubMed=9813312;
RA   Itoh A., Miyabayashi T., Ohno M., Sakano S.;
RT   "Cloning and expressions of three mammalian homologues of drosophila
RT   slit suggest possible roles for slit in the formation and maintenance
RT   of the nervous system.";
RL   Brain Res. Mol. Brain Res. 62:175-186(1998).
DR   EMBL; AB017167; BAA35184.1; -.
DR   HSSP; P00743; 1CCF.
DR   GO; GO:0005615; C:extracellular space; TAS.
DR   GO; GO:0005515; F:protein binding; TAS.
DR   GO; GO:0007411; P:axon guidance; TAS.
DR   GO; GO:0007420; P:brain development; TAS.
DR   GO; GO:0007275; P:development; TAS.
DR   GO; GO:0007399; P:neurogenesis; TAS.
DR   InterPro; IPR000152; Asx_hydroxyl_S.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR006207; Cys_knot_C.
DR   InterPro; IPR000742; EGF_2.
DR   InterPro; IPR001881; EGF_Ca.
DR   InterPro; IPR006209; EGF_like.
DR   InterPro; IPR003645; FolN.
DR   InterPro; IPR002049; Laminin_EGF.
DR   InterPro; IPR001791; Laminin_G.
DR   InterPro; IPR001611; LRR.
DR   InterPro; IPR000483; LRR_Cterm.
DR   InterPro; IPR000372; LRR_Nterm.
DR   InterPro; IPR003591; LRR_typ.
DR   Pfam; PF00008; EGF; 9.
DR   Pfam; PF00054; laminin_G; 1.
DR   Pfam; PF00560; LRR; 19.
DR   Pfam; PF01463; LRRCT; 4.
DR   Pfam; PF01462; LRRNT; 4.
DR   PRINTS; PR00011; EGFLAMININ.
DR   SMART; SM00041; CT; 1.
DR   SMART; SM00179; EGF_CA; 2.
DR   SMART; SM00274; FOLN; 3.
DR   SMART; SM00282; LamG; 1.
DR   SMART; SM00082; LRRCT; 4.
DR   SMART; SM00013; LRRNT; 4.
DR   SMART; SM00369; LRR_TYP; 10.
DR   PROSITE; PS00010; ASX_HYDROXYL; 2.
DR   PROSITE; PS01185; CTCK_1; 1.
DR   PROSITE; PS01225; CTCK_2; 1.
DR   PROSITE; PS00022; EGF_1; 9.
DR   PROSITE; PS01186; EGF_2; 8.
DR   PROSITE; PS01187; EGF_CA; 2.
DR   PROSITE; PS50025; LAM_G_DOMAIN; 1.
KW   EGF-like domain.
SQ   SEQUENCE   1534 AA;  167952 MW;  8954EF8EA4DAEBA1 CRC64;

  Query Match             18.6%;  Score 292;  DB 4;  Length 1534;
  Best Local Similarity   28.0%;  Pred. No. 4.5e-14;
  Matches   82;  Conservative   48;  Mismatches   91;  Indels   72;  Gaps   14;

Qy           1 CERQPCQHGATCMPAGEYEFQCLCRDGFKGDLCE------HEENPCQLREPCLHGGTC-- 52
               |    ||:|| ||      : ||| :|: | |||        ::||:  | | :|  |
Db        1087 CRDHRCQNGAQCMDEVN-SYSCLCAEGYSGQLCEIPPHLPAPKSPCEGTE-CQNGANCVD 1144

Qy          53 QGTR--CLCLPGFSGPRCQQGSGHGIAESDWHLEGSGGNDAPGQYGAYFHDDGFLAFPGH 110
               || |  | ||||| || |::       : | :|:              | |
Db        1145 QGNRPVCQCLPGFGGPECEKLLSVNFVDRDTYLQ--------------FTD--------- 1181

Qy         111 VFSRSLPEVPE-TIELEVRTSTASGLLLWQGVEVGEAGQGKDFISLGLQDGHLVFRYQLG 169
                    |   |   | |:| |:  :|:||: |          | |:: |  ||:   |  |
Db        1182 -----LQNWPRANITLQVSTAEDNGILLYNG--------DNDHIAVELYQGHVRVSYDPG 1228

Qy         170 S-GEARLVSEDPINDGEWHRVTALREGRRGSIQVDGEELVSGRSPGPNVAVNAKGSVYIG 228
               |   : : | : ||||::| |  :   :  :: :||   ::  : | :  :|::  :|:|
Db        1229 SYPSSAIYSAETINDGQFHTVELVAFDQMVNLSIDGGSPMTMDNFGKHYTLNSEAPLYVG 1288

Qy         229 GAP-DVAT--------LTGGRFSSGITGCVKNLVLHS---------ARPGAPP 263
               | | || :        | |    :|  ||::|| :::          :||  |
Db        1289 GMPVDVNSAAFRLWQILNG----TGFHGCIRNLYINNELQDFTKTQMKPGVVP 1337

RESULT 15
Q9UIL7
ID   Q9UIL7      PRELIMINARY;      PRT;  1618 AA.
AC   Q9UIL7;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   MEGF4 (Fragment).
GN   MEGF4.
OS   Homo sapiens (Human).
```

```
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Brain;
RX   MEDLINE=98360089; PubMed=9693030;
RA   Nakayama M., Nakajima D., Nagase T., Nomura N., Seki N., Ohara O.;
RT   "Identification of high-molecular-weight proteins with multiple EGF-
RT   like motifs by motif-trap screening.";
RL   Genomics 51:27-34(1998).
DR   EMBL; AB011537; BAA32465.3; -.
DR   HSSP; P00743; 1CCF.
DR   Genew; HGNC:11085; SLIT1.
DR   GO; GO:0005509; F:calcium ion binding; NAS.
DR   InterPro; IPR000152; Asx_hydroxyl_S.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR006207; Cys_knot_C.
DR   InterPro; IPR000742; EGF_2.
DR   InterPro; IPR001881; EGF_Ca.
DR   InterPro; IPR001438; EGF_II.
DR   InterPro; IPR006209; EGF_like.
DR   InterPro; IPR006210; IEGF.
DR   InterPro; IPR002049; Laminin_EGF.
DR   InterPro; IPR001791; Laminin_G.
DR   InterPro; IPR001611; LRR.
DR   InterPro; IPR000483; LRR_Cterm.
DR   InterPro; IPR000372; LRR_Nterm.
DR   InterPro; IPR003591; LRR_typ.
DR   Pfam; PF00008; EGF; 9.
DR   Pfam; PF00054; laminin_G; 1.
DR   Pfam; PF00560; LRR; 19.
DR   Pfam; PF01463; LRRCT; 4.
DR   Pfam; PF01462; LRRNT; 4.
DR   PRINTS; PR00010; EGFBLOOD.
DR   PRINTS; PR00011; EGFLAMININ.
DR   SMART; SM00041; CT; 1.
DR   SMART; SM00181; EGF; 9.
DR   SMART; SM00179; EGF_CA; 9.
DR   SMART; SM00282; LamG; 1.
DR   SMART; SM00082; LRRCT; 4.
DR   SMART; SM00013; LRRNT; 4.
DR   SMART; SM00369; LRR_TYP; 18.
DR   PROSITE; PS00010; ASX_HYDROXYL; 2.
DR   PROSITE; PS01185; CTCK_1; 1.
DR   PROSITE; PS01225; CTCK_2; 1.
DR   PROSITE; PS00022; EGF_1; 9.
DR   PROSITE; PS01186; EGF_2; 8.
DR   PROSITE; PS01187; EGF_CA; 2.
DR   PROSITE; PS50025; LAM_G_DOMAIN; 1.
KW   EGF-like domain.
FT   NON_TER       1      1
SQ   SEQUENCE   1618 AA;  176732 MW;  BF62DF0478B034E0 CRC64;

  Query Match             18.6%;  Score 292;  DB 4;  Length 1618;
  Best Local Similarity   28.0%;  Pred. No. 4.8e-14;
  Matches   82;  Conservative   48;  Mismatches   91;  Indels   72;  Gaps   14;

Qy          1 CERQPCQHGATCMPAGEYEFQCLCRDGFKGDLCE------HEENPCQLREPCLHGGTC-- 52
              |    ||:|| ||      : ||| :|: | |||        ::||:  | | :|  |
Db       1171 CRDHRCQNGAQCMDEVN-SYSCLCAEGYSGQLCEIPPHLPAPKSPCEGTE-CQNGANCVD 1228

Qy         53 QGTR--CLCLPGFSGPRCQQGSGHGIAESDWHLEGSGGNDAPGQYGAYFHDDGFLAFPGH 110
              || |  | ||||| || |::       : | :|:              | |
Db       1229 QGNRPVCQCLPGFGGPECEKLLSVNFVDRDTYLQ--------------FTD--------- 1265

Qy        111 VFSRSLPEVPE-TIELEVRTSTASGLLLWQGVEVGEAGQGKDFISLGLQDGHLVFRYQLG 169
                   |   |   | |:| |:  :|:||: |          | |:: |  ||:   |  |
Db       1266 -----LQNWPRANITLQVSTAEDNGILLYNG--------DNDHIAVELYQGHVRVSYDPG 1312

Qy        170 S-GEARLVSEDPINDGEWHRVTALREGRRGSIQVDGEELVSGRSPGPNVAVNAKGSVYIG 228
              |   : : | : ||||::| |  :   :  :: :||   ::  : | :  :|::  :|:|
Db       1313 SYPSSAIYSAETINDGQFHTVELVAFDQMVNLSIDGGSPMTMDNFGKHYTLNSEAPLYVG 1372

Qy        229 GAP-DVAT--------LTGGRFSSGITGCVKNLVLHS---------ARPGAPP 263
              | | || :        | |    :|  ||::|| :::          :||  |
Db       1373 GMPVDVNSAAFRLWQILNG----TGFHGCIRNLYINNELQDFTKTQMKPGVVP 1421

Search completed: March  9, 2004, 17:21:43
Job time : 25.8707 secs
```

```
                          GenCore version 5.1.6
                 Copyright (c) 1993 - 2004  Compugen Ltd.

OM protein - protein search, using sw model

Run on:          March  9, 2004, 17:19:48 ; Search time 34.5822 Seconds
                                            (without alignments)
                                            2320.373 Million cell updates/sec

Title:           US-10-006-011A-9
Perfect score:   284
Sequence:        1 CERQPCQHGATCMPAGEYEF..........QPLDLQHRAQAGANTRPCPS 284

Scoring table:   OLIGO
                 Gapop 60.0 , Gapext 60.0

Searched:        1586107 seqs, 282547505 residues

Word size :       0

Total number of hits satisfying chosen parameters:      1586107

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :        A_Geneseq_29Jan04:*
                 1:  geneseqp1980s:*
                 2:  geneseqp1990s:*
                 3:  geneseqp2000s:*
                 4:  geneseqp2001s:*
                 5:  geneseqp2002s:*
                 6:  geneseqp2003as:*
                 7:  geneseqp2003bs:*
                 8:  geneseqp2004s:*

      Pred. No. is the number of results predicted by chance to have a
      score greater than or equal to the score of the result being printed,
      and is derived by analysis of the total score distribution.

                                SUMMARIES
                %
Result          Query
  No.   Score   Match Length DB  ID                    Description
 -------------------------------------------------------------------------
     1    284   100.0    4391  6  AAE34390              Aae34390 Human per
     2    196    69.0    4393  4  AAB31889              Aab31889 Amino aci
     3    195    68.7     195  4  AAB31890              Aab31890 Amino aci
     4    164    57.7    4436  4  ABG23265              Abg23265 Novel hum
     5     18     6.3      18  4  ABB55911              Abb55911 Vascular
     6     14     4.9      14  4  ABB55908              Abb55908 Vascular
     7     14     4.9      14  4  ABB55905              Abb55905 Vascular
     8     10     3.5      10  4  ABB55910              Abb55910 Vascular
     9      9     3.2      15  4  ABB55909              Abb55909 Vascular
    10      9     3.2     159  6  ABM65732              Abm65732 Propionib
    11      9     3.2    2531  7  ADE63713              Ade63713 Rat Prote
    12      9     3.2    2531  7  ADE63705              Ade63705 Rat Prote
    13      9     3.2    2531  7  ADE63709              Ade63709 Rat Prote
    14      9     3.2    2531  7  ADE63701              Ade63701 Rat Prote
    15      9     3.2    3319  4  ABB70376              Abb70376 Drosophil
    16      8     2.8       8  4  ABB55907              Abb55907 Vascular
    17      8     2.8      14  4  ABB55906              Abb55906 Vascular
    18      8     2.8      54  4  AAU40544              Aau40544 Propionib
    19      8     2.8      54  6  ABM37063              Abm37063 Propionib
    20      8     2.8      80  4  AAM16081              Aam16081 Peptide #
    21      8     2.8      80  4  ABB35070              Abb35070 Peptide #
    22      8     2.8      80  4  AAM28574              Aam28574 Peptide #
    23      8     2.8      80  4  ABB29890              Abb29890 Peptide #
    24      8     2.8      80  4  ABB20486              Abb20486 Protein #
    25      8     2.8      80  4  AAM68258              Aam68258 Human bon
    26      8     2.8      80  4  AAM55885              Aam55885 Human bra
    27      8     2.8      80  4  ABG49912              Abg49912 Human liv
    28      8     2.8      80  4  AAM03808              Aam03808 Peptide #
    29      8     2.8      80  5  ABG37793              Abg37793 Human pep
    30      8     2.8     110  5  ABB53124              Abb53124 Human ORF
    31      8     2.8     144  5  ABB53123              Abb53123 Human ORF
    32      8     2.8     144  5  ABB53141              Abb53141 Human ORF
    33      8     2.8     161  4  ABG20412              Abg20412 Novel hum
    34      8     2.8     248  3  AAB40530              Aab40530 Human ORF
    35      8     2.8     248  5  ABP05276              Abp05276 Human ORF
    36      8     2.8     445  4  AAU42626              Aau42626 Propionib
    37      8     2.8     445  6  ABM39145              Abm39145 Propionib
    38      8     2.8     494  6  ABU18011              Abu18011 Protein e
    39      8     2.8    1130  2  AAR13436              Aar13436 Merosin M
    40      8     2.8    1130  2  AAR71729              Aar71729 Merosin m
    41      8     2.8    1257  2  AAR46627              Aar46627 Neurocan
    42      8     2.8    1307  7  AAE39974              Aae39974 Human CRU
    43      8     2.8    1321  6  ABP97200              Abp97200 Tumour-as
    44      8     2.8    1388  5  ABP70118              Abp70118 Human NOV
    45      8     2.8    1388  5  ABP70117              Abp70117 Human NOV

                                ALIGNMENTS


RESULT 1
AAE34390
ID   AAE34390 standard; protein; 4391 AA.
XX
AC   AAE34390;
XX
DT   14-MAY-2003  (first entry)
XX
DE   Human perlecan protein.
XX
KW   Human; diagnosis; osteoarthritis; rheumatoid arthritis; perlecan.
XX
OS   Homo sapiens.
XX
PN   WO200295415-A2.
XX
PD   28-NOV-2002.
XX
PF   22-MAY-2002; 2002WO-EP005612.
XX
PR   23-MAY-2001; 2001GB-00012626.
XX
PA   (OSTE-) OSTEOMETER BIO TECH AS.
XX
PI   Christgau S,  Henriksen DB,  Cloos PAC;
XX
DR   WPI; 2003-140389/13.
XX
PT   An assay for the diagnosis or assessment of the severity of
PT   osteoarthritis or rheumatoid arthritis comprising detecting an isomerized
PT   or optically inverted protein in a sample.
XX
PS   Disclosure; Page 46-67; 106pp; English.
XX
CC   The invention relates to an assay for the diagnosis or assessment of the
CC   severity of osteoarthritis or rheumatoid arthritis. The assay involves
CC   measuring (in a biological sample) the amount or presence of an
CC   isomerised or optically inverted protein or one or more isomerised or
CC   optically inverted fragments from proteins such as perlecan, biglycan,
CC   decorin, fibrillin-1 or protocadherin. The assay is useful for the
CC   diagnosis or assessment of the severity of osteoarthritis or rheumatoid
CC   arthritis. The present sequence is human perlecan protein
XX
SQ   Sequence 4391 AA;

  Query Match             100.0%;  Score 284;  DB 6;  Length 4391;
  Best Local Similarity   100.0%;  Pred. No. 6.9e-268;
  Matches  284;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;
```

```
Qy          1 CERQPCQHGATCMPAGEYEFQCLCRDGFKGDLCEHEENPCQLREPCLHGGTCQGTRCLCL 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4108 CERQPCQHGATCMPAGEYEFQCLCRDGFKGDLCEHEENPCQLREPCLHGGTCQGTRCLCL 4167

Qy         61 PGFSGPRCQQGSGHGIAESDWHLEGSGGNDAPGQYGAYFHDDGFLAFPGHVFSRSLPEVP 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4168 PGFSGPRCQQGSGHGIAESDWHLEGSGGNDAPGQYGAYFHDDGFLAFPGHVFSRSLPEVP 4227

Qy        121 ETIELEVRTSTASGLLLWQGVEVGEAGQGKDFISLGLQDGHLVFRYQLGSGEARLVSEDP 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4228 ETIELEVRTSTASGLLLWQGVEVGEAGQGKDFISLGLQDGHLVFRYQLGSGEARLVSEDP 4287

Qy        181 INDGEWHRVTALREGRRGSIQVDGEELVSGRSPGPNVAVNAKGSVYIGGAPDVATLTGGR 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4288 INDGEWHRVTALREGRRGSIQVDGEELVSGRSPGPNVAVNAKGSVYIGGAPDVATLTGGR 4347

Qy        241 FSSGITGCVKNLVLHSARPGAPPPQPLDLQHRAQAGANTRPCPS 284
              ||||||||||||||||||||||||||||||||||||||||||||
Db       4348 FSSGITGCVKNLVLHSARPGAPPPQPLDLQHRAQAGANTRPCPS 4391

RESULT 2
AAB31889
ID   AAB31889 standard; protein; 4393 AA.
XX
AC   AAB31889;
XX
DT   15-MAY-2001  (first entry)
XX
DE   Amino acid sequence of a human protein.
XX
KW   Human; perlecan; retinol-binding plasma protein; calgranulin B; vaccine;
KW   ganglioside GM2 activator; saposin B; degenerative disease; glial cell;
KW   neurological disease; auto-immune disease; multiple sclerosis; toxicity;
KW   Alzheimer's disease; Parkinson's disease; amyotrophic lateral sclerosis;
KW   rheumatoid polyarthritis; lupus erythematosus; gene therapy.
XX
OS   Homo sapiens.
XX
PN   WO200105422-A2.
XX
PD   25-JAN-2001.
XX
PF   17-JUL-2000; 2000WO-FR002057.
XX
PR   15-JUL-1999;   99FR-00009372.
XX
PA   (INMR ) BIOMERIEUX STELHYS.
XX
PI   Roecklin D,  Kolbe H,  Charles M,  Malcus C,  Santoro L,  Perron H;
XX
DR   WPI; 2001-159475/16.
XX
PT   Detecting, preventing and treating degenerative, neurological and
PT   autoimmune diseases, particularly multiple sclerosis, using specified
PT   polypeptides or related nucleic acid or ligand.
XX
PS   Claim 1; Page 138-152; 209pp; French.
XX
CC   The present sequence represents a human protein, which is used in the
CC   method of the invention. The specification describes a method which uses
CC   at least one polypeptide or polynucleotide sequence belonging to the
CC   perlecan, precursor of the retinol-binding plasma protein, precursor of
CC   the ganglioside GM2 activator, calgranulin B or saposin B protein
CC   families. The method is used for detecting, preventing or treating a
CC   degenerative, neurological and/or auto-immune disease. The
CC   polynucleotides and polypeptides are used for diagnosis, prognosis,
CC   prevention and treatment of multiple sclerosis (in its various forms and
CC   phases). They may also be useful in cases of e.g. Alzheimer's and
CC   Parkinson's diseases, amyotrophic lateral sclerosis, rheumatoid
CC   polyarthritis and lupus erythematosus, including use as vaccines and in
CC   gene therapy (expression of sense or antisense sequences). They can also
CC   be used to assess efficacy of potential therapeutic agents, particularly
CC   compounds that reduce or inhibit toxicity towards glial cells
XX
SQ   Sequence 4393 AA;

  Query Match             69.0%;  Score 196;  DB 4;  Length 4393;
  Best Local Similarity   100.0%;  Pred. No. 7.9e-182;
  Matches  196;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         29 KGDLCEHEENPCQLREPCLHGGTCQGTRCLCLPGFSGPRCQQGSGHGIAESDWHLEGSGG 88
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4138 KGDLCEHEENPCQLREPCLHGGTCQGTRCLCLPGFSGPRCQQGSGHGIAESDWHLEGSGG 4197

Qy         89 NDAPGQYGAYFHDDGFLAFPGHVFSRSLPEVPETIELEVRTSTASGLLLWQGVEVGEAGQ 148
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4198 NDAPGQYGAYFHDDGFLAFPGHVFSRSLPEVPETIELEVRTSTASGLLLWQGVEVGEAGQ 4257

Qy        149 GKDFISLGLQDGHLVFRYQLGSGEARLVSEDPINDGEWHRVTALREGRRGSIQVDGEELV 208
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4258 GKDFISLGLQDGHLVFRYQLGSGEARLVSEDPINDGEWHRVTALREGRRGSIQVDGEELV 4317

Qy        209 SGRSPGPNVAVNAKGS 224
              ||||||||||||||||
Db       4318 SGRSPGPNVAVNAKGS 4333

RESULT 3
AAB31890
ID   AAB31890 standard; protein; 195 AA.
XX
AC   AAB31890;
XX
DT   15-MAY-2001  (first entry)
XX
DE   Amino acid sequence of the C-terminal of the human perlecan protein.
XX
KW   Human; perlecan; retinol-binding plasma protein; calgranulin B; vaccine;
KW   ganglioside GM2 activator; saposin B; degenerative disease; glial cell;
KW   neurological disease; auto-immune disease; multiple sclerosis; toxicity;
KW   Alzheimer's disease; Parkinson's disease; amyotrophic lateral sclerosis;
KW   rheumatoid polyarthritis; lupus erythematosus; gene therapy.
XX
OS   Homo sapiens.
XX
PN   WO200105422-A2.
XX
PD   25-JAN-2001.
XX
PF   17-JUL-2000; 2000WO-FR002057.
XX
PR   15-JUL-1999;   99FR-00009372.
XX
PA   (INMR ) BIOMERIEUX STELHYS.
XX
PI   Roecklin D,  Kolbe H,  Charles M,  Malcus C,  Santoro L,  Perron H;
XX
DR   WPI; 2001-159475/16.
DR   N-PSDB; AAF54728.
XX
PT   Detecting, preventing and treating degenerative, neurological and
PT   autoimmune diseases, particularly multiple sclerosis, using specified
PT   polypeptides or related nucleic acid or ligand.
XX
PS   Claim 1; Page 152-153; 209pp; French.
XX
CC   The present sequence represents a human polypeptide, which is used in the
CC   method of the invention. The specification describes a method which uses
CC   at least one polypeptide or polynucleotide sequence belonging to the
CC   perlecan, precursor of the retinol-binding plasma protein, precursor of
CC   the ganglioside GM2 activator, calgranulin B or saposin B protein
CC   families. The method is used for detecting, preventing or treating a
CC   degenerative, neurological and/or auto-immune disease. The
CC   polynucleotides and polypeptides are used for diagnosis, prognosis,
```

```
CC   prevention and treatment of multiple sclerosis (in its various forms and
CC   phases). They may also be useful in cases of e.g. Alzheimer's and
CC   Parkinson's diseases, amyotrophic lateral sclerosis, rheumatoid
CC   polyarthritis and lupus erythematosus, including use as vaccines and in
CC   gene therapy (expression of sense or antisense sequences). They can also
CC   be used to assess efficacy of potential therapeutic agents, particularly
CC   compounds that reduce or inhibit toxicity towards glial cells
XX
SQ   Sequence 195 AA;

  Query Match             68.7%;  Score 195;  DB 4;  Length 195;
  Best Local Similarity   100.0%;  Pred. No. 4.2e-182;
  Matches  195;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          90 DAPGQYGAYFHDDGFLAFPGHVFSRSLPEVPETIELEVRTSTASGLLLWQGVEVGEAGQG 149
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db           1 DAPGQYGAYFHDDGFLAFPGHVFSRSLPEVPETIELEVRTSTASGLLLWQGVEVGEAGQG 60

Qy         150 KDFISLGLQDGHLVFRYQLGSGEARLVSEDPINDGEWHRVTALREGRRGSIQVDGEELVS 209
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          61 KDFISLGLQDGHLVFRYQLGSGEARLVSEDPINDGEWHRVTALREGRRGSIQVDGEELVS 120

Qy         210 GRSPGPNVAVNAKGSVYIGGAPDVATLTGGRFSSGITGCVKNLVLHSARPGAPPPQPLDL 269
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         121 GRSPGPNVAVNAKGSVYIGGAPDVATLTGGRFSSGITGCVKNLVLHSARPGAPPPQPLDL 180

Qy         270 QHRAQAGANTRPCPS 284
               |||||||||||||||
Db         181 QHRAQAGANTRPCPS 195


RESULT 4
ABG23265
ID   ABG23265 standard; protein; 4436 AA.
XX
AC   ABG23265;
XX
DT   18-FEB-2002  (first entry)
XX
DE   Novel human diagnostic protein #23256.
XX
KW   Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW   food supplement; medical imaging; diagnostic; genetic disorder.
XX
OS   Homo sapiens.
XX
PN   WO200175067-A2.
XX
PD   11-OCT-2001.
XX
PF   30-MAR-2001; 2001WO-US008631.
XX
PR   31-MAR-2000; 2000US-00540217.
PR   23-AUG-2000; 2000US-00649167.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Drmanac RT,  Liu C,  Tang YT;
XX
DR   WPI; 2001-639362/73.
DR   N-PSDB; AAS87452.
XX
PT   New isolated polynucleotide and encoded polypeptides, useful in
PT   diagnostics, forensics, gene mapping, identification of mutations
PT   responsible for genetic disorders or other traits and to assess
PT   biodiversity.
XX
PS   Claim 20; SEQ ID NO 53624; 103pp; English.
XX
CC   The invention relates to isolated polynucleotide (I) and polypeptide (II)
CC   sequences. (I) is useful as hybridisation probes, polymerase chain
CC   reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
CC   and in recombinant production of (II). The polynucleotides are also used
CC   in diagnostics as expressed sequence tags for identifying expressed
CC   genes. (I) is useful in gene therapy techniques to restore normal
CC   activity of (II) or to treat disease states involving (II). (II) is
CC   useful for generating antibodies against it, detecting or quantitating a
CC   polypeptide in tissue, as molecular weight markers and as a food
CC   supplement. (II) and its binding partners are useful in medical imaging
CC   of sites expressing (II). (I) and (II) are useful for treating disorders
CC   involving aberrant protein expression or biological activity. The
CC   polypeptide and polynucleotide sequences have applications in
CC   diagnostics, forensics, gene mapping, identification of mutations
CC   responsible for genetic disorders or other traits to assess biodiversity
CC   and to produce other types of data and products dependent on DNA and
CC   amino acid sequences. ABG00010-ABG30377 represent novel human diagnostic
CC   amino acid sequences of the invention. Note: The sequence data for this
CC   patent did not appear in the printed specification, but was obtained in
CC   electronic format directly from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 4436 AA;

  Query Match             57.7%;  Score 164;  DB 4;  Length 4436;
  Best Local Similarity   100.0%;  Pred. No. 1.6e-150;
  Matches  164;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy           1 CERQPCQHGATCMPAGEYEFQCLCRDGFKGDLCEHEENPCQLREPCLHGGTCQGTRCLCL 60
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        4143 CERQPCQHGATCMPAGEYEFQCLCRDGFKGDLCEHEENPCQLREPCLHGGTCQGTRCLCL 4202

Qy          61 PGFSGPRCQQGSGHGIAESDWHLEGSGGNDAPGQYGAYFHDDGFLAFPGHVFSRSLPEVP 120
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        4203 PGFSGPRCQQGSGHGIAESDWHLEGSGGNDAPGQYGAYFHDDGFLAFPGHVFSRSLPEVP 4262

Qy         121 ETIELEVRTSTASGLLLWQGVEVGEAGQGKDFISLGLQDGHLVF 164
               ||||||||||||||||||||||||||||||||||||||||||||
Db        4263 ETIELEVRTSTASGLLLWQGVEVGEAGQGKDFISLGLQDGHLVF 4306


RESULT 5
ABB55911
ID   ABB55911 standard; peptide; 18 AA.
XX
AC   ABB55911;
XX
DT   15-FEB-2002  (first entry)
XX
DE   Vascular dementia-associated protein isoform (VPI) 111.
XX
KW   Vascular Dementia; VD; VD-associated protein isoform; VPI; screening;
KW   diagnosis; prognosis; gene therapy.
XX
OS   Homo sapiens.
XX
PN   WO200169261-A2.
XX
PD   20-SEP-2001.
XX
PF   14-MAR-2001; 2001WO-GB001106.
XX
PR   15-MAR-2000; 2000GB-00006285.
PR   24-NOV-2000; 2000GB-00028734.
PR   28-NOV-2000; 2000US-00724391.
XX
PA   (OXFO-) OXFORD GLYCOSCIENCES UK LTD.
XX
PI   Herath HMAC,  Parekh RB,  Rohlff C;
XX
DR   WPI; 2001-557937/62.
XX
PT   Screening, diagnosis or prognosis of vascular dementia (VD), useful for
PT   determining stage of VD and monitoring the effect of VD therapy,
PT   comprises analyzing body fluid by 2-dimensional electrophoresis for
PT   features correlated with VD.
XX
```

```
PS   Claim 6; Page 32; 151pp; English.
XX
CC   The invention relates to screening, diagnosis or prognosis of Vascular
CC   Dementia (VD) in a subject comprising analysing body fluid from the
CC   subject by 2-dimensional (2-D) electrophoresis to generate a 2-D array of
CC   features containing at least one chosen feature whose relative abundance
CC   correlates with the presence, absence, stage or severity of VD or
CC   predicts the onset or course of VD, especially detecting in a sample of
CC   cerebrospinal fluid (CSF) from the subject one of 223 VD-associated
CC   protein isoforms (VPIs) (ABB55801-ABB56295) as fully defined in the
CC   specification. Detecting VD-associated features and VPI is useful for the
CC   screening, diagnosis or prognosis of VD, for determining the stage or
CC   severity of VD, for identifying a subject at risk of VD or for monitoring
CC   the effect of therapy administered to a subject having VD. Nucleic acids
CC   encoding a VPI or inhibiting the function of a VPI are useful for the
CC   treatment of VD and for gene therapy
XX
SQ   Sequence 18 AA;

  Query Match             6.3%;  Score 18;  DB 4;  Length 18;
  Best Local Similarity   100.0%;  Pred. No. 5.8e-10;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        223 GSVYIGGAPDVATLTGGR 240
              ||||||||||||||||||
Db          1 GSVYIGGAPDVATLTGGR 18

RESULT 6
ABB55908
ID   ABB55908 standard; peptide; 14 AA.
XX
AC   ABB55908;
XX
DT   15-FEB-2002  (first entry)
XX
DE   Vascular dementia-associated protein isoform (VPI) 108.
XX
KW   Vascular Dementia; VD; VD-associated protein isoform; VPI; screening;
KW   diagnosis; prognosis; gene therapy.
XX
OS   Homo sapiens.
XX
PN   WO200169261-A2.
XX
PD   20-SEP-2001.
XX
PF   14-MAR-2001; 2001WO-GB001106.
XX
PR   15-MAR-2000; 2000GB-00006285.
PR   24-NOV-2000; 2000GB-00028734.
PR   28-NOV-2000; 2000US-00724391.
XX
PA   (OXFO-) OXFORD GLYCOSCIENCES UK LTD.
XX
PI   Herath HMAC,  Parekh RB,  Rohlff C;
XX
DR   WPI; 2001-557937/62.
XX
PT   Screening, diagnosis or prognosis of vascular dementia (VD), useful for
PT   determining stage of VD and monitoring the effect of VD therapy,
PT   comprises analyzing body fluid by 2-dimensional electrophoresis for
PT   features correlated with VD.
XX
PS   Claim 6; Page 32; 151pp; English.
XX
CC   The invention relates to screening, diagnosis or prognosis of Vascular
CC   Dementia (VD) in a subject comprising analysing body fluid from the
CC   subject by 2-dimensional (2-D) electrophoresis to generate a 2-D array of
CC   features containing at least one chosen feature whose relative abundance
CC   correlates with the presence, absence, stage or severity of VD or
CC   predicts the onset or course of VD, especially detecting in a sample of
CC   cerebrospinal fluid (CSF) from the subject one of 223 VD-associated
```

```
CC   protein isoforms (VPIs) (ABB55801-ABB56295) as fully defined in the
CC   specification. Detecting VD-associated features and VPI is useful for the
CC   screening, diagnosis or prognosis of VD, for determining the stage or
CC   severity of VD, for identifying a subject at risk of VD or for monitoring
CC   the effect of therapy administered to a subject having VD. Nucleic acids
CC   encoding a VPI or inhibiting the function of a VPI are useful for the
CC   treatment of VD and for gene therapy
XX
SQ   Sequence 14 AA;

  Query Match             4.9%;  Score 14;  DB 4;  Length 14;
  Best Local Similarity   100.0%;  Pred. No. 3.8e-06;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        175 LVSEDPINDGEWHR 188
              ||||||||||||||
Db          1 LVSEDPINDGEWHR 14

RESULT 7
ABB55905
ID   ABB55905 standard; peptide; 14 AA.
XX
AC   ABB55905;
XX
DT   15-FEB-2002  (first entry)
XX
DE   Vascular dementia-associated protein isoform (VPI) 105.
XX
KW   Vascular Dementia; VD; VD-associated protein isoform; VPI; screening;
KW   diagnosis; prognosis; gene therapy.
XX
OS   Homo sapiens.
XX
PN   WO200169261-A2.
XX
PD   20-SEP-2001.
XX
PF   14-MAR-2001; 2001WO-GB001106.
XX
PR   15-MAR-2000; 2000GB-00006285.
PR   24-NOV-2000; 2000GB-00028734.
PR   28-NOV-2000; 2000US-00724391.
XX
PA   (OXFO-) OXFORD GLYCOSCIENCES UK LTD.
XX
PI   Herath HMAC,  Parekh RB,  Rohlff C;
XX
DR   WPI; 2001-557937/62.
XX
PT   Screening, diagnosis or prognosis of vascular dementia (VD), useful for
PT   determining stage of VD and monitoring the effect of VD therapy,
PT   comprises analyzing body fluid by 2-dimensional electrophoresis for
PT   features correlated with VD.
XX
PS   Claim 6; Page 32; 151pp; English.
XX
CC   The invention relates to screening, diagnosis or prognosis of Vascular
CC   Dementia (VD) in a subject comprising analysing body fluid from the
CC   subject by 2-dimensional (2-D) electrophoresis to generate a 2-D array of
CC   features containing at least one chosen feature whose relative abundance
CC   correlates with the presence, absence, stage or severity of VD or
CC   predicts the onset or course of VD, especially detecting in a sample of
CC   cerebrospinal fluid (CSF) from the subject one of 223 VD-associated
CC   protein isoforms (VPIs) (ABB55801-ABB56295) as fully defined in the
CC   specification. Detecting VD-associated features and VPI is useful for the
CC   screening, diagnosis or prognosis of VD, for determining the stage or
CC   severity of VD, for identifying a subject at risk of VD or for monitoring
CC   the effect of therapy administered to a subject having VD. Nucleic acids
CC   encoding a VPI or inhibiting the function of a VPI are useful for the
CC   treatment of VD and for gene therapy
XX
SQ   Sequence 14 AA;
```

```
  Query Match             4.9%;  Score 14;  DB 4;  Length 14;
  Best Local Similarity   100.0%;  Pred. No. 3.8e-06;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        198 GSIQVDGEELVSGR 211
              ||||||||||||||
Db          1 GSIQVDGEELVSGR 14

RESULT 8
ABB55910
ID   ABB55910 standard; peptide; 10 AA.
XX
AC   ABB55910;
XX
DT   15-FEB-2002  (first entry)
XX
DE   Vascular dementia-associated protein isoform (VPI) 110.
XX
KW   Vascular Dementia; VD; VD-associated protein isoform; VPI; screening;
KW   diagnosis; prognosis; gene therapy.
XX
OS   Homo sapiens.
XX
PN   WO200169261-A2.
XX
PD   20-SEP-2001.
XX
PF   14-MAR-2001; 2001WO-GB001106.
XX
PR   15-MAR-2000; 2000GB-00006285.
PR   24-NOV-2000; 2000GB-00028734.
PR   28-NOV-2000; 2000US-00724391.
XX
PA   (OXFO-) OXFORD GLYCOSCIENCES UK LTD.
XX
PI   Herath HMAC,  Parekh RB,  Rohlff C;
XX
DR   WPI; 2001-557937/62.
XX
PT   Screening, diagnosis or prognosis of vascular dementia (VD), useful for
PT   determining stage of VD and monitoring the effect of VD therapy,
PT   comprises analyzing body fluid by 2-dimensional electrophoresis for
PT   features correlated with VD.
XX
PS   Claim 6; Page 32; 151pp; English.
XX
CC   The invention relates to screening, diagnosis or prognosis of Vascular
CC   Dementia (VD) in a subject comprising analysing body fluid from the
CC   subject by 2-dimensional (2-D) electrophoresis to generate a 2-D array of
CC   features containing at least one chosen feature whose relative abundance
CC   correlates with the presence, absence, stage or severity of VD or
CC   predicts the onset or course of VD, especially detecting in a sample of
CC   cerebrospinal fluid (CSF) from the subject one of 223 VD-associated
CC   protein isoforms (VPIs) (ABB55801-ABB56295) as fully defined in the
CC   specification. Detecting VD-associated features and VPI is useful for the
CC   screening, diagnosis or prognosis of VD, for determining the stage or
CC   severity of VD, for identifying a subject at risk of VD or for monitoring
CC   the effect of therapy administered to a subject having VD. Nucleic acids
CC   encoding a VPI or inhibiting the function of a VPI are useful for the
CC   treatment of VD and for gene therapy
XX
SQ   Sequence 10 AA;

  Query Match             3.5%;  Score 10;  DB 4;  Length 10;
  Best Local Similarity   100.0%;  Pred. No. 0.022;
  Matches   10;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        241 FSSGITGCVK 250
              ||||||||||
Db          1 FSSGITGCVK 10
```

```
RESULT 9
ABB55909
ID   ABB55909 standard; peptide; 15 AA.
XX
AC   ABB55909;
XX
DT   15-FEB-2002  (first entry)
XX
DE   Vascular dementia-associated protein isoform (VPI) 109.
XX
KW   Vascular Dementia; VD; VD-associated protein isoform; VPI; screening;
KW   diagnosis; prognosis; gene therapy.
XX
OS   Homo sapiens.
XX
PN   WO200169261-A2.
XX
PD   20-SEP-2001.
XX
PF   14-MAR-2001; 2001WO-GB001106.
XX
PR   15-MAR-2000; 2000GB-00006285.
PR   24-NOV-2000; 2000GB-00028734.
PR   28-NOV-2000; 2000US-00724391.
XX
PA   (OXFO-) OXFORD GLYCOSCIENCES UK LTD.
XX
PI   Herath HMAC,  Parekh RB,  Rohlff C;
XX
DR   WPI; 2001-557937/62.
XX
PT   Screening, diagnosis or prognosis of vascular dementia (VD), useful for
PT   determining stage of VD and monitoring the effect of VD therapy,
PT   comprises analyzing body fluid by 2-dimensional electrophoresis for
PT   features correlated with VD.
XX
PS   Claim 6; Page 32; 151pp; English.
XX
CC   The invention relates to screening, diagnosis or prognosis of Vascular
CC   Dementia (VD) in a subject comprising analysing body fluid from the
CC   subject by 2-dimensional (2-D) electrophoresis to generate a 2-D array of
CC   features containing at least one chosen feature whose relative abundance
CC   correlates with the presence, absence, stage or severity of VD or
CC   predicts the onset or course of VD, especially detecting in a sample of
CC   cerebrospinal fluid (CSF) from the subject one of 223 VD-associated
CC   protein isoforms (VPIs) (ABB55801-ABB56295) as fully defined in the
CC   specification. Detecting VD-associated features and VPI is useful for the
CC   screening, diagnosis or prognosis of VD, for determining the stage or
CC   severity of VD, for identifying a subject at risk of VD or for monitoring
CC   the effect of therapy administered to a subject having VD. Nucleic acids
CC   encoding a VPI or inhibiting the function of a VPI are useful for the
CC   treatment of VD and for gene therapy
XX
SQ   Sequence 15 AA;

  Query Match             3.2%;  Score 9;  DB 4;  Length 15;
  Best Local Similarity   100.0%;  Pred. No. 0.31;
  Matches    9;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        166 YQLGSGEAR 174
              |||||||||
Db          1 YQLGSGEAR 9

RESULT 10
ABM65732
ID   ABM65732 standard; protein; 159 AA.
XX
AC   ABM65732;
XX
DT   20-OCT-2003  (first entry)
XX
```

```
DE   Propionibacterium acnes immunogenic polypeptide #30408.
XX
KW   Acne vulgaris; antiseborrhoeic; dermatological; antibacterial;
KW   immunostimulant; immune response; vaccine; immunogenic.
XX
OS   Propionibacterium acnes.
XX
PN   WO2003033515-A1.
XX
PD   24-APR-2003.
XX
PF   11-OCT-2002; 2002WO-US032727.
XX
PR   15-OCT-2001; 2001US-00978825.
XX
PA   (CORI-) CORIXA CORP.
XX
PI   Mitcham JL,  Skeiky YAW,  Persing DH,  Bhatia A,  Maisonneuve JL;
PI   Zhang Y,  Wang S,  Jen S,  Lodes MJ,  Benson DR,  Jones R,  Carter D;
PI   Barth B,  Vallieve-Douglass J;
XX
DR   WPI; 2003-381789/36.
XX
PT   New Propionibacterium acnes polypeptides and polynucleotides encoding the
PT   polypeptide, useful for diagnosing, preventing or treating acne vulgaris,
PT   or for stimulating an immune response specific for a P. acnes protein.
XX
PS   Claim 7; SEQ ID NO 30408; 1481pp; English.
XX
CC   The invention relates to an isolated polynucleotide (ACF64435-ACF64733)
CC   encoding a Propionibacterium acnes protein. The invention also relates to
CC   polypeptides encoded by the polynucleotides (ABM35624-ABM64536) and to
CC   immunogenic fragments of P. acnes polypeptides. The invention
CC   additionally encompasses expression vectors and host cells comprising a
CC   polynucleotide of the invention; antibodies against polypeptides of the
CC   invention; fusion proteins comprising a polypeptide of the invention; a
CC   method for stimulating an immune response specific for a P. acnes
CC   polypeptide and an isolated T cell population comprising T cells prepared
CC   via this method; a vaccine composition (comprising P. acnes polypeptides,
CC   polynucleotides, antibodies, fusion proteins, T cell populations, or
CC   antigen-presenting cells that express the polypeptide); a method and kit
CC   for detecting or determining the presence or absence of P. acnes in a
CC   patient; and a method for inhibiting the development of P. acnes in a
CC   patient. The P. acnes polypeptides, polynucleotides, antibodies, fusion
CC   proteins, T cell populations or antigen-presenting cells that express the
CC   polypeptides are useful for diagnosing, preventing or treating acne
CC   vulgaris, or for stimulating an immune response specific for a P. acnes
CC   protein. The polynucleotides can also be used as probes or primers for
CC   nucleic acid hybridisation. The vaccine composition is useful for the
CC   stimulation of an immune response against P. acnes, or for treating acne,
CC   and the kit is useful for performing a diagnostic assay. The present
CC   sequence represents a specifically claimed P. acnes polypeptide which is
CC   thought to contain an immunogenic region. Note: The sequence data for
CC   this patent did not form part of the printed specification, but was
CC   obtained in electronic format directly from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 159 AA;

  Query Match             3.2%;  Score 9;  DB 6;  Length 159;
  Best Local Similarity   100.0%;  Pred. No. 2.7;
  Matches    9;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        127 VRTSTASGL 135
              |||||||||
Db         31 VRTSTASGL 39


RESULT 11
ADE63713
ID   ADE63713 standard; protein; 2531 AA.
XX
AC   ADE63713;
```

```
XX
DT   29-JAN-2004  (first entry)
XX
DE   Rat Protein CAA40667, SEQ ID NO 9657.
XX
KW   Rat; pain; neuronal tissue; gene therapy; spinal segmental nerve injury;
KW   chronic constriction injury; CCI; spared nerve injury; SNI; Chung.
XX
OS   Rattus norvegicus.
XX
PN   WO2003016475-A2.
XX
PD   27-FEB-2003.
XX
PF   14-AUG-2002; 2002WO-US025765.
XX
PR   14-AUG-2001; 2001US-0312147P.
PR   01-NOV-2001; 2001US-0346382P.
PR   26-NOV-2001; 2001US-0333347P.
XX
PA   (GEHO ) GEN HOSPITAL CORP.
PA   (FARB ) BAYER AG.
XX
PI   Woolf C,  D'urso D,  Befort K,  Costigan M;
XX
DR   WPI; 2003-268312/26.
DR   GENBANK; CAA40667.
XX
PT   New composition comprising two or more isolated polypeptides, useful for
PT   preparing a medicament for treating pain in an animal.
XX
PS   Claim 1; Page; 1017pp; English.
XX
CC   The invention discloses a composition comprising two or more isolated rat
CC   or human polynucleotides or a polynucleotide which represents a fragment,
CC   derivative or allelic variation of the nucleic acid sequence. Also
CC   claimed are a vector comprising the novel polynucleotide, a host cell
CC   comprising the vector, a method for identifying a nucleotide sequence
CC   which is differentially regulated in an animal subjected to pain and a
CC   kit to perform the method, an array, a method for identifying an agent
CC   that increases or decreases the expression of the polynucleotide sequence
CC   that is differentially expressed in neuronal tissue of a first animal
CC   subjected to pain, a method for identifying a compound which regulates
CC   the expression of a polynucleotide sequence which is differentially
CC   expressed in an animal subjected to pain, a method for identifying a
CC   compound that regulates the activity of one or more of the
CC   polynucleotides, a method for producing a pharmaceutical composition, a
CC   method for identifying a compound or small molecule that regulates the
CC   activity in an animal of one or more of the polypeptides given in the
CC   specification, a method for identifying a compound useful in treating
CC   pain and a pharmaceutical composition comprising the one or more
CC   polypeptides or their antibodies. The polynucleotide or the compound that
CC   modulates its activity is useful for preparing a medicament for treating
CC   pain (e.g. spinal segmental nerve injury (Chung), chronic constriction
CC   injury (CCI) and spared nerve injury (SNI)) in an animal (e.g. gene
CC   therapy). The sequence presented is a rat protein (shown in Table 2 of
CC   the specification) which is differentially expressed during pain. Note:
CC   The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic form directly from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 2531 AA;

  Query Match             3.2%;  Score 9;  DB 7;  Length 2531;
  Best Local Similarity   100.0%;  Pred. No. 35;
  Matches    9;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         45 PCLHGGTCQ 53
              |||||||||
Db       1029 PCLHGGTCQ 1037


RESULT 12
```

```
ADE63705
ID   ADE63705 standard; protein; 2531 AA.
XX
AC   ADE63705;
XX
DT   29-JAN-2004  (first entry)
XX
DE   Rat Protein CAA40667, SEQ ID NO 9649.
XX
KW   Rat; pain; neuronal tissue; gene therapy; spinal segmental nerve injury;
KW   chronic constriction injury; CCI; spared nerve injury; SNI; Chung.
XX
OS   Rattus norvegicus.
XX
PN   WO2003016475-A2.
XX
PD   27-FEB-2003.
XX
PF   14-AUG-2002; 2002WO-US025765.
XX
PR   14-AUG-2001; 2001US-0312147P.
PR   01-NOV-2001; 2001US-0346382P.
PR   26-NOV-2001; 2001US-0333347P.
XX
PA   (GEHO ) GEN HOSPITAL CORP.
PA   (FARB ) BAYER AG.
XX
PI   Woolf C,  D'urso D,  Befort K,  Costigan M;
XX
DR   WPI; 2003-268312/26.
DR   GENBANK; CAA40667.
XX
PT   New composition comprising two or more isolated polypeptides, useful for
PT   preparing a medicament for treating pain in an animal.
XX
PS   Claim 1; Page; 1017pp; English.
XX
CC   The invention discloses a composition comprising two or more isolated rat
CC   or human polynucleotides or a polynucleotide which represents a fragment,
CC   derivative or allelic variation of the nucleic acid sequence. Also
CC   claimed are a vector comprising the novel polynucleotide, a host cell
CC   comprising the vector, a method for identifying a nucleotide sequence
CC   which is differentially regulated in an animal subjected to pain and a
CC   kit to perform the method, an array, a method for identifying an agent
CC   that increases or decreases the expression of the polynucleotide sequence
CC   that is differentially expressed in neuronal tissue of a first animal
CC   subjected to pain, a method for identifying a compound which regulates
CC   the expression of a polynucleotide sequence which is differentially
CC   expressed in an animal subjected to pain, a method for identifying a
CC   compound that regulates the activity of one or more of the
CC   polynucleotides, a method for producing a pharmaceutical composition, a
CC   method for identifying a compound or small molecule that regulates the
CC   activity in an animal of one or more of the polypeptides given in the
CC   specification, a method for identifying a compound useful in treating
CC   pain and a pharmaceutical composition comprising the one or more
CC   polypeptides or their antibodies. The polynucleotide or the compound that
CC   modulates its activity is useful for preparing a medicament for treating
CC   pain (e.g. spinal segmental nerve injury (Chung), chronic constriction
CC   injury (CCI) and spared nerve injury (SNI)) in an animal (e.g. gene
CC   therapy). The sequence presented is a rat protein (shown in Table 2 of
CC   the specification) which is differentially expressed during pain. Note:
CC   The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic form directly from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 2531 AA;

  Query Match             3.2%;  Score 9;  DB 7;  Length 2531;
  Best Local Similarity   100.0%;  Pred. No. 35;
  Matches    9;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         45 PCLHGGTCQ 53
              |||||||||
Db       1029 PCLHGGTCQ 1037

RESULT 13
ADE63709
ID   ADE63709 standard; protein; 2531 AA.
XX
AC   ADE63709;
XX
DT   29-JAN-2004  (first entry)
XX
DE   Rat Protein CAA40667, SEQ ID NO 9653.
XX
KW   Rat; pain; neuronal tissue; gene therapy; spinal segmental nerve injury;
KW   chronic constriction injury; CCI; spared nerve injury; SNI; Chung.
XX
OS   Rattus norvegicus.
XX
PN   WO2003016475-A2.
XX
PD   27-FEB-2003.
XX
PF   14-AUG-2002; 2002WO-US025765.
XX
PR   14-AUG-2001; 2001US-0312147P.
PR   01-NOV-2001; 2001US-0346382P.
PR   26-NOV-2001; 2001US-0333347P.
XX
PA   (GEHO ) GEN HOSPITAL CORP.
PA   (FARB ) BAYER AG.
XX
PI   Woolf C,  D'urso D,  Befort K,  Costigan M;
XX
DR   WPI; 2003-268312/26.
DR   GENBANK; CAA40667.
XX
PT   New composition comprising two or more isolated polypeptides, useful for
PT   preparing a medicament for treating pain in an animal.
XX
PS   Claim 1; Page; 1017pp; English.
XX
CC   The invention discloses a composition comprising two or more isolated rat
CC   or human polynucleotides or a polynucleotide which represents a fragment,
CC   derivative or allelic variation of the nucleic acid sequence. Also
CC   claimed are a vector comprising the novel polynucleotide, a host cell
CC   comprising the vector, a method for identifying a nucleotide sequence
CC   which is differentially regulated in an animal subjected to pain and a
CC   kit to perform the method, an array, a method for identifying an agent
CC   that increases or decreases the expression of the polynucleotide sequence
CC   that is differentially expressed in neuronal tissue of a first animal
CC   subjected to pain, a method for identifying a compound which regulates
CC   the expression of a polynucleotide sequence which is differentially
CC   expressed in an animal subjected to pain, a method for identifying a
CC   compound that regulates the activity of one or more of the
CC   polynucleotides, a method for producing a pharmaceutical composition, a
CC   method for identifying a compound or small molecule that regulates the
CC   activity in an animal of one or more of the polypeptides given in the
CC   specification, a method for identifying a compound useful in treating
CC   pain and a pharmaceutical composition comprising the one or more
CC   polypeptides or their antibodies. The polynucleotide or the compound that
CC   modulates its activity is useful for preparing a medicament for treating
CC   pain (e.g. spinal segmental nerve injury (Chung), chronic constriction
CC   injury (CCI) and spared nerve injury (SNI)) in an animal (e.g. gene
CC   therapy). The sequence presented is a rat protein (shown in Table 2 of
CC   the specification) which is differentially expressed during pain. Note:
CC   The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic form directly from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 2531 AA;

  Query Match             3.2%;  Score 9;  DB 7;  Length 2531;
  Best Local Similarity   100.0%;  Pred. No. 35;
```

```
  Matches     9;  Conservative     0;  Mismatches     0;  Indels     0;  Gaps     0;

Qy         45 PCLHGGTCQ 53
              |||||||||
Db       1029 PCLHGGTCQ 1037

RESULT 14
ADE63701
ID   ADE63701 standard; protein; 2531 AA.
XX
AC   ADE63701;
XX
DT   29-JAN-2004  (first entry)
XX
DE   Rat Protein CAA40667, SEQ ID NO 9645.
XX
KW   Rat; pain; neuronal tissue; gene therapy; spinal segmental nerve injury;
KW   chronic constriction injury; CCI; spared nerve injury; SNI; Chung.
XX
OS   Rattus norvegicus.
XX
PN   WO2003016475-A2.
XX
PD   27-FEB-2003.
XX
PF   14-AUG-2002; 2002WO-US025765.
XX
PR   14-AUG-2001; 2001US-0312147P.
PR   01-NOV-2001; 2001US-0346382P.
PR   26-NOV-2001; 2001US-0333347P.
XX
PA   (GEHO ) GEN HOSPITAL CORP.
PA   (FARB ) BAYER AG.
XX
PI   Woolf C,  D'urso D,  Befort K,  Costigan M;
XX
DR   WPI; 2003-268312/26.
DR   GENBANK; CAA40667.
XX
PT   New composition comprising two or more isolated polypeptides, useful for
PT   preparing a medicament for treating pain in an animal.
XX
PS   Claim 1; Page; 1017pp; English.
XX
CC   The invention discloses a composition comprising two or more isolated rat
CC   or human polynucleotides or a polynucleotide which represents a fragment,
CC   derivative or allelic variation of the nucleic acid sequence. Also
CC   claimed are a vector comprising the novel polynucleotide, a host cell
CC   comprising the vector, a method for identifying a nucleotide sequence
CC   which is differentially regulated in an animal subjected to pain and a
CC   kit to perform the method, an array, a method for identifying an agent
CC   that increases or decreases the expression of the polynucleotide sequence
CC   that is differentially expressed in neuronal tissue of a first animal
CC   subjected to pain, a method for identifying a compound which regulates
CC   the expression of a polynucleotide sequence which is differentially
CC   expressed in an animal subjected to pain, a method for identifying a
CC   compound that regulates the activity of one or more of the
CC   polynucleotides, a method for producing a pharmaceutical composition, a
CC   method for identifying a compound or small molecule that regulates the
CC   activity in an animal of one or more of the polypeptides given in the
CC   specification, a method for identifying a compound useful in treating
CC   pain and a pharmaceutical composition comprising the one or more
CC   polypeptides or their antibodies. The polynucleotide or the compound that
CC   modulates its activity is useful for preparing a medicament for treating
CC   pain (e.g. spinal segmental nerve injury (Chung), chronic constriction
CC   injury (CCI) and spared nerve injury (SNI)) in an animal (e.g. gene
CC   therapy). The sequence presented is a rat protein (shown in Table 2 of
CC   the specification) which is differentially expressed during pain. Note:
CC   The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic form directly from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences.
XX
```

```
SQ   Sequence 2531 AA;

  Query Match             3.2%;  Score 9;  DB 7;  Length 2531;
  Best Local Similarity   100.0%;  Pred. No. 35;
  Matches    9;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         45 PCLHGGTCQ 53
              |||||||||
Db     , 1029 PCLHGGTCQ 1037

RESULT 15
ABB70376
ID   ABB70376 standard; protein; 3319 AA.
XX
AC   ABB70376;
XX
DT   26-MAR-2002  (first entry)
XX
DE   Drosophila melanogaster polypeptide SEQ ID NO 37920.
XX
KW   Drosophila; developmental biology; cell signalling; insecticide;
KW   pharmaceutical.
XX
OS   Drosophila melanogaster.
XX
PN   WO200171042-A2.
XX
PD   27-SEP-2001.
XX
PF   23-MAR-2001; 2001WO-US009231.
XX
PR   23-MAR-2000; 2000US-0191637P.
PR   11-JUL-2000; 2000US-00614150.
XX
PA   (PEKE ) PE CORP NY.
XX
PI   Venter JC,  Adams M,  Li PWD,  Myers EW;
XX
DR   WPI; 2001-656860/75.
DR   N-PSDB; ABL14479.
XX
PT   New isolated nucleic acid detection reagent for detecting 1000 or more
PT   genes from Drosophila and for elucidating cell signaling and cell-cell
PT   interactions.
XX
PS   Disclosure; SEQ ID NO 37920; 21pp + Sequence Listing; English.
XX
CC   The invention relates to an isolated nucleic acid detection reagent
CC   capable of detecting 1000 or more genes from Drosophila. The invention is
CC   useful in developmental biology and in elucidating cell signalling and
CC   cell-cell interactions in higher eukaryotes for the development of
CC   insecticides, therapeutics and pharmaceutical drugs. The invention
CC   discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC   sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-
CC   ABB72072). The sequence data for this patent did not form part of the
CC   printed specification, but was obtained in electronic format directly
CC   from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 3319 AA;

  Query Match             3.2%;  Score 9;  DB 4;  Length 3319;
  Best Local Similarity   100.0%;  Pred. No. 45;
  Matches    9;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        182 NDGEWHRVT 190
              |||||||||
Db       3033 NDGEWHRVT 3041


Search completed: March  9, 2004, 17:27:54
Job time : 35.5822 secs
```

```
                        GenCore version 5.1.6
                 Copyright (c) 1993 - 2004  Compugen Ltd.

OM protein - protein search, using sw model

Run on:          March  9, 2004, 17:25:24 ; Search time 10.8957 Seconds
                                            (without alignments)
                                            1345.642 Million cell updates/sec

Title:           US-10-006-011A-9
Perfect score:   284
Sequence:        1 CERQPCQHGATCMPAGEYEF..........QPLDLQHRAQAGANTRPCPS 284

Scoring table:   OLIGO
                 Gapop 60.0 , Gapext 60.0

Searched:        389414 seqs, 51625971 residues

Word size :      0

Total number of hits satisfying chosen parameters:      389414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :        Issued_Patents_AA:*
                 1:  /cgn2_6/ptodata/2/iaa/5A_COMB.pep:*
                 2:  /cgn2_6/ptodata/2/iaa/5B_COMB.pep:*
                 3:  /cgn2_6/ptodata/2/iaa/6A_COMB.pep:*
                 4:  /cgn2_6/ptodata/2/iaa/6B_COMB.pep:*
                 5:  /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep:*
                 6:  /cgn2_6/ptodata/2/iaa/backfiles1.pep:*

       Pred. No. is the number of results predicted by chance to have a
       score greater than or equal to the score of the result being printed,
       and is derived by analysis of the total score distribution.

                                     SUMMARIES
                    %
Result            Query
   No.    Score   Match Length DB  ID                              Description
-------------------------------------------------------------------------------
     1       9     3.2    1940  2  US-08-644-271-30                 Sequence 30, Appl
     2       9     3.2    1940  4  US-09-077-955-34                 Sequence 34, Appl
     3       8     2.8    1015  1  US-08-537-210A-1                 Sequence 1, Appli
     4       8     2.8    1015  3  US-09-113-825-1                  Sequence 1, Appli
     5       8     2.8    1130  2  US-08-460-309-2                  Sequence 2, Appli
     6       8     2.8    1130  2  US-08-125-077-2                  Sequence 2, Appli
     7       8     2.8    1130  6  5444158-2                       Patent No. 5444158
     8       8     2.8    1257  1  US-08-340-428B-49                Sequence 49, Appl
     9       8     2.8    2321  4  US-09-230-652-2                  Sequence 2, Appli
    10       8     2.8    2471  1  US-08-185-432-16                 Sequence 16, Appl
    11       8     2.8    2471  1  US-08-083-590A-19                Sequence 19, Appl
    12       8     2.8    2471  3  US-08-532-384-19                 Sequence 19, Appl
    13       8     2.8    2471  4  US-08-899-232-1                  Sequence 1, Appli
    14       8     2.8    3084  4  US-09-562-702A-12                Sequence 12, Appl
    15       8     2.8    3088  4  US-09-562-702A-8                 Sequence 8, Appli
    16       8     2.8    3089  4  US-09-562-702A-4                 Sequence 4, Appli
    17       8     2.8    3106  4  US-09-562-702A-10                Sequence 10, Appl
    18       8     2.8    3110  4  US-09-562-702A-2                 Sequence 2, Appli
    19       8     2.8    3110  4  US-09-562-702A-6                 Sequence 6, Appli
    20       8     2.8    3110  4  US-09-561-709B-7                 Sequence 7, Appli
    21       8     2.8    3111  2  US-08-460-309-4                  Sequence 4, Appli
    22       8     2.8    3111  2  US-08-125-077-4                  Sequence 4, Appli
    23       7     2.5      77  4  US-09-252-991A-18061             Sequence 18061, A
    24       7     2.5     165  4  US-09-252-991A-27759             Sequence 27759, A
    25       7     2.5     181  4  US-09-252-991A-23483             Sequence 23483, A
    26       7     2.5     186  4  US-09-107-532A-6672              Sequence 6672, Ap
    27       7     2.5     216  1  US-08-315-695-20                 Sequence 20, Appl
    28       7     2.5     237  4  US-09-252-991A-28116             Sequence 28116, A
    29       7     2.5     240  4  US-09-570-856B-16                Sequence 16, Appl
    30       7     2.5     244  4  US-09-252-991A-18465             Sequence 18465, A
    31       7     2.5     248  2  US-08-755-559-1                  Sequence 1, Appli
    32       7     2.5     248  3  US-09-210-474-1                  Sequence 1, Appli
    33       7     2.5     248  4  US-09-539-774-1                  Sequence 1, Appli
    34       7     2.5     273  4  US-09-252-991A-28324             Sequence 28324, A
    35       7     2.5     290  4  US-09-252-991A-27099             Sequence 27099, A
    36       7     2.5     349  4  US-09-162-524-3                  Sequence 3, Appli
    37       7     2.5     352  4  US-09-252-991A-19989             Sequence 19989, A
    38       7     2.5     372  4  US-09-252-991A-20108             Sequence 20108, A
    39       7     2.5     381  3  US-08-911-853-27                 Sequence 27, Appl
    40       7     2.5     381  3  US-09-479-409-27                 Sequence 27, Appl
    41       7     2.5     381  4  US-09-479-453-27                 Sequence 27, Appl
    42       7     2.5     384  4  US-09-252-991A-24086             Sequence 24086, A
    43       7     2.5     410  4  US-09-252-991A-31487             Sequence 31487, A
    44       7     2.5     519  4  US-09-328-352-7681               Sequence 7681, Ap
    45       7     2.5     728  4  US-09-107-532A-4518              Sequence 4518, Ap

                                  ALIGNMENTS

RESULT 1
US-08-644-271-30
; Sequence 30, Application US/08644271
; Patent No. 5814478
;  GENERAL INFORMATION:
;    APPLICANT:  Valenzuela, et al.
;    TITLE OF INVENTION:  NOVEL TYROSINE KINASE RECEPTORS
;    TITLE OF INVENTION:  AND LIGANDS
;    NUMBER OF SEQUENCES:  32
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Regeneron Pharmaceuticals, Inc.
;      STREET:  777 Old Saw Mill Road
;      CITY:  Tarrytown
;      STATE:  NY
;      COUNTRY:  USA
;      ZIP:  10591
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/644,271
;      FILING DATE:  10-MAY-1996
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  USSN 60/008,657
;      FILING DATE:  15-DEC-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Cobert, Robert J
;      REGISTRATION NUMBER:  36,108
;      REFERENCE/DOCKET NUMBER:  REG 195A
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  914-345-7400
;      TELEFAX:  914-345-7721
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  30:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1940 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
;    FEATURE:
;      NAME/KEY:  Rat Agrin
;      LOCATION:  1...1940
;      OTHER INFORMATION:
US-08-644-271-30

  Query Match             3.2%;  Score 9;  DB 2;  Length 1940;
```

```
  Best Local Similarity   100.0%;  Pred. No. 5;
  Matches    9;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         45 PCLHGGTCQ 53
              |||||||||
Db       1228 PCLHGGTCQ 1236


RESULT 2
US-09-077-955-34
; Sequence 34, Application US/09077955A
; Patent No. 6413740
; GENERAL INFORMATION:
;  APPLICANT: Valenzuela et al., David M.
;  TITLE OF INVENTION: NOVEL TYROSINE KINASE RECEPTORS AND LIGANDS
;  FILE REFERENCE: REG195-B-PCT-US
;  CURRENT APPLICATION NUMBER: US/09/077,955A
;  CURRENT FILING DATE: 1998-09-10
;  EARLIER APPLICATION NUMBER: PCT/US96/20696
;  EARLIER FILING DATE: 1996-12-13
;  EARLIER APPLICATION NUMBER: 08/644,271
;  EARLIER FILING DATE: 1996-05-10
;  EARLIER APPLICATION NUMBER: 60/008,657
;  EARLIER FILING DATE: 1995-12-15
;  NUMBER OF SEQ ID NOS: 36
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 34
;   LENGTH: 1940
;   TYPE: PRT
;   ORGANISM: Rattus sp.
US-09-077-955-34

  Query Match             3.2%;  Score 9;  DB 4;  Length 1940;
  Best Local Similarity   100.0%;  Pred. No. 5;
  Matches    9;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         45 PCLHGGTCQ 53
              |||||||||
Db       1228 PCLHGGTCQ 1236


RESULT 3
US-08-537-210A-1
; Sequence 1, Application US/08537210A
; Patent No. 5780300
;  GENERAL INFORMATION:
;    APPLICANT:  Artavanis-Tsakonas, Spyridon
;    APPLICANT:  Fortini, Mark
;    APPLICANT:  Matsuno, Kenji
;    TITLE OF INVENTION:  MANIPULATION OF NON-TERMINALLY
;    TITLE OF INVENTION:  DIFFERENTIATED CELLS USING THE NOTCH PATHWAY
;    NUMBER OF SEQUENCES:  4
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Pennie & Edmonds
;      STREET:  1155 Avenue of the Americas
;      CITY:  New York
;      STATE:  NY
;      COUNTRY:  USA
;      ZIP:  10036/2711
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/537,210A
;      FILING DATE:  29-SEP-1995
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Misrock, S. Leslie
;      REGISTRATION NUMBER:  18,872
;      REFERENCE/DOCKET NUMBER:  7326-027
;    TELECOMMUNICATION INFORMATION:
```

```
;      TELEPHONE:  212-790-9090
;      TELEFAX:  212-869-8864
;      TELEX:  66141 PENNIE
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1015 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  protein
;    FEATURE:
;      NAME/KEY:  hum N (Human No. 5780300ch 2)
;      LOCATION:  1155...2169
;      OTHER INFORMATION:  Highly conserved ankyrin repeat
;      OTHER INFORMATION:  region of No. 5780300ch
US-08-537-210A-1

  Query Match             2.8%;  Score 8;  DB 1;  Length 1015;
  Best Local Similarity   100.0%;  Pred. No. 27;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          5 PCQHGATC 12
              ||||||||
Db          3 PCQHGATC 10


RESULT 4
US-09-113-825-1
; Sequence 1, Application US/09113825
; Patent No. 6149902
;  GENERAL INFORMATION:
;    APPLICANT:  Artavanis-Tsakonas, Spyridon
;    APPLICANT:  Fortini, Mark
;    APPLICANT:  Matsuno, Kenji
;    TITLE OF INVENTION:  MANIPULATION OF NON-TERMINALLY
;    TITLE OF INVENTION:  DIFFERENTIATED CELLS USING THE NOTCH PATHWAY
;    NUMBER OF SEQUENCES:  4
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Pennie & Edmonds
;      STREET:  1155 Avenue of the Americas
;      CITY:  New York
;      STATE:  NY
;      COUNTRY:  USA
;      ZIP:  10036/2711
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/113,825
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/537,210
;      FILING DATE:  29-SEP-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Misrock, S. Leslie
;      REGISTRATION NUMBER:  18,872
;      REFERENCE/DOCKET NUMBER:  7326-027
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  212-790-9090
;      TELEFAX:  212-869-8864
;      TELEX:  66141 PENNIE
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1015 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  protein
;    FEATURE:
;      NAME/KEY:  hum N (Human No. 6149902ch 2)
```

```
;      LOCATION:  1155...2169
;      OTHER INFORMATION:  Highly conserved ankyrin repeat
;      OTHER INFORMATION:  region of No. 6149902ch
US-09-113-825-1

  Query Match              2.8%;  Score 8;  DB 3;  Length 1015;
  Best Local Similarity   100.0%;  Pred. No. 27;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          5 PCQHGATC 12
              ||||||||
Db          3 PCQHGATC 10


RESULT 5
US-08-460-309-2
; Sequence 2, Application US/08460309
; Patent No. 5837496
;  GENERAL INFORMATION:
;    APPLICANT:  Engvall, Eva
;    APPLICANT:  Leivo, Ilmo
;    TITLE OF INVENTION:  Nucleic Acids Encoding Merosin, Merosin
;    TITLE OF INVENTION:  Fragments and Uses Thereof
;    NUMBER OF SEQUENCES:  23
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Campbell and Flores
;      STREET:  4370 La Jolla Village Drive, Suite 700
;      CITY:  San Diego
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  92122
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/460,309
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/125,077
;      FILING DATE:  22-SEP-1993
;      APPLICATION NUMBER:  US PCT/US 94/10730
;      FILING DATE:  21-SEP-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/472,319
;      FILING DATE:  30-JAN-1990
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/919,951
;      FILING DATE:  27-JUL-1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Campbell, Cathryn A.
;      REGISTRATION NUMBER:  31,815
;      REFERENCE/DOCKET NUMBER:  P-LA 9721
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (619) 535-9001
;      TELEFAX:  (619) 535-8949
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1130 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-460-309-2

  Query Match              2.8%;  Score 8;  DB 2;  Length 1130;
  Best Local Similarity   100.0%;  Pred. No. 29;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        122 TIELEVRT 129
              ||||||||
Db        808 TIELEVRT 815


RESULT 6
US-08-125-077-2
; Sequence 2, Application US/08125077
; Patent No. 5872231
; Patent No. 5872231 5840863
;  GENERAL INFORMATION:
;    APPLICANT:  Engvall, Eva
;    APPLICANT:  Leivo, Ilmo
;    TITLE OF INVENTION:  Nucleic Acids Encoding Merosin, Merosin
;    TITLE OF INVENTION:  Fragments and Uses Thereof
;    NUMBER OF SEQUENCES:  23
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Campbell and Flores
;      STREET:  4370 La Jolla Village Drive, Suite 700
;      CITY:  San Diego
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  92122
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/125,077
;      FILING DATE:  22-SEP-1993
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US PCT/US 94/10730
;      FILING DATE:  21-SEP-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/472,319
;      FILING DATE:  30-JAN-1990
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/919,951
;      FILING DATE:  27-JUL-1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Campbell, Cathryn A.
;      REGISTRATION NUMBER:  31,815
;      REFERENCE/DOCKET NUMBER:  P-LA 9721
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (619) 535-9001
;      TELEFAX:  (619) 535-8949
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1130 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-125-077-2

  Query Match              2.8%;  Score 8;  DB 2;  Length 1130;
  Best Local Similarity   100.0%;  Pred. No. 29;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        122 TIELEVRT 129
              ||||||||
Db        808 TIELEVRT 815


RESULT 7
5444158-2
;Patent No. 5444158
;    APPLICANT: ENGVALL, EVA;SANES, JOSHUA
;    TITLE OF INVENTION: MEROSIN, NUCLEIC ACIDS ENCODING,
;FRAGMENTS AND USES THEREOF
;    NUMBER OF SEQUENCES: 4
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/87,642
;      FILING DATE: 08-JUL-1993
;    PRIOR APPLICATION DATA:
```

```
;      APPLICATION NUMBER: 587,689
;      FILING DATE: 24-SEP-1990
;      APPLICATION NUMBER: 472,319
;      FILING DATE: 30-JAN-1990
;SEQ ID NO:2:
;      LENGTH: 1130
5444158-2

  Query Match             2.8%;  Score 8;  DB 6;  Length 1130;
  Best Local Similarity   100.0%;  Pred. No. 29;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        122 TIELEVRT 129
              ||||||||
Db        808 TIELEVRT 815

RESULT 8
US-08-340-428B-49
; Sequence 49, Application US/08340428B
; Patent No. 5648465
;  GENERAL INFORMATION:
;    APPLICANT:  MARGOLIS, Richard U.
;    APPLICANT:  RAUCH, Uwe
;    APPLICANT:  MARGOLIS, Renee K.
;    TITLE OF INVENTION:  CLONING, EXPRESSION AND USES FOR A
;    TITLE OF INVENTION:  NEUROCAN AS A CHONDROITIN SULFATE PROTEOGLYCAN
;    NUMBER OF SEQUENCES:  49
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Browdy and Neimark
;      STREET:  419 Seventh Street, N.W.
;      CITY:  Washington
;      STATE:  D.C.
;      COUNTRY:  U.S.A.
;      ZIP:  20004
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/340,428B
;      FILING DATE:  14 No. 5648465ember 1994
;      CLASSIFICATION:  514
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/922,911
;      FILING DATE:  03 August 1992
;      CLASSIFICATION:  514
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Browdy, Roger L.
;      REGISTRATION NUMBER:  25,618
;      REFERENCE/DOCKET NUMBER:  Margolis=1A
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  202-628-5197
;      TELEFAX:  202-737-3528
;  INFORMATION FOR SEQ ID NO:  49:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1257 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
US-08-340-428B-49

  Query Match             2.8%;  Score 8;  DB 1;  Length 1257;
  Best Local Similarity   100.0%;  Pred. No. 33;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         45 PCLHGGTC 52
              ||||||||
Db        957 PCLHGGTC 964
```

```
RESULT 9
US-09-230-652-2
; Sequence 2, Application US/09230652A
; Patent No. 6537775
; GENERAL INFORMATION:
;  APPLICANT: Tournier-Lasserve, Elisabeth
;  APPLICANT: Joutel, Anne
;  APPLICANT: Bousser, Marie-Germaine
;  APPLICANT: Bach, Jean-Francois
;  TITLE OF INVENTION: GENE INVOLVED IN CADASIL, METHOD OF DIAGNOSIS AND
;  TITLE OF INVENTION: THERAPEUTIC APPLICATION
;  FILE REFERENCE: 03715.0048-00000
;  CURRENT APPLICATION NUMBER: US/09/230,652A
;  CURRENT FILING DATE: 1999-05-17
;  EARLIER APPLICATION NUMBER: FR 96 09733
;  EARLIER FILING DATE: 1996-08-01
;  EARLIER APPLICATION NUMBER: FR 97 04680
;  EARLIER FILING DATE: 1997-04-16
;  EARLIER APPLICATION NUMBER: PCT/FR97/01433
;  EARLIER FILING DATE: 1997-07-31
;  NUMBER OF SEQ ID NOS: 163
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
;   LENGTH: 2321
;   TYPE: PRT
;   ORGANISM: Homo sapiens
;   FEATURE:
;   OTHER INFORMATION: human ADNc No. 6537775ch 3
US-09-230-652-2

  Query Match             2.8%;  Score 8;  DB 4;  Length 2321;
  Best Local Similarity   100.0%;  Pred. No. 58;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         62 GFSGPRCQ 69
              ||||||||
Db       1237 GFSGPRCQ 1244


RESULT 10
US-08-185-432-16
; Sequence 16, Application US/08185432
; Patent No. 5750652
;  GENERAL INFORMATION:
;    APPLICANT:  Artavanis-Tsakonas, Spyridon
;    APPLICANT:  Busseau, Isabelle
;    APPLICANT:  Diederich, Robert J.
;    APPLICANT:  Xu, Tian
;    APPLICANT:  Matsuno, Kenji
;    TITLE OF INVENTION:  DELTEX PROTEINS, NUCLEIC ACIDS, AND
;    TITLE OF INVENTION:  ANTIBODIES, AND RELATED METHODS AND COMPOSITIONS
;    NUMBER OF SEQUENCES:  23
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  PENNIE & EDMONDS
;      STREET:  1155 Avenue of the Americas
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  U.S.A.
;      ZIP:  10036-2711
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/185,432
;      FILING DATE:  21-JAN-1994
;      CLASSIFICATION:  530
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Misrock, S. Leslie
;      REGISTRATION NUMBER:  18,872
;      REFERENCE/DOCKET NUMBER:  7326-006
;    TELECOMMUNICATION INFORMATION:
```

```
;      TELEPHONE:  (212) 790-9090
;      TELEFAX:  (212) 869-8864/9741
;      TELEX:  66141 PENNIE
;  INFORMATION FOR SEQ ID NO:  16:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2471 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  protein
US-08-185-432-16

  Query Match             2.8%;  Score 8;  DB 1;  Length 2471;
  Best Local Similarity   100.0%;  Pred. No. 62;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          5 PCQHGATC 12
              ||||||||
Db       1157 PCQHGATC 1164


RESULT 11
US-08-083-590A-19
; Sequence 19, Application US/08083590A
; Patent No. 5786158
;  GENERAL INFORMATION:
;    APPLICANT:  Artavanis-Tsakonas, S. et al.
;    TITLE OF INVENTION:  Therapeutic And Diagnostic Methods
;    TITLE OF INVENTION:  And Compositions Based On No. 5786158ch Proteins And
;    TITLE OF INVENTION:  Nucleic Acids
;    NUMBER OF SEQUENCES:  21
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Pennie & Edmonds
;      STREET:  1155 Avenue of the Americas
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  U.S.A.
;      ZIP:  10036
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/083,590A
;      FILING DATE:  25-JUN-1993
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Misrock, S. Leslie
;      REGISTRATION NUMBER:  18,872
;      REFERENCE/DOCKET NUMBER:  7326-015
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  212 790-9090
;      TELEFAX:  212 8698864/9741
;      TELEX:  66141 PENNIE
;  INFORMATION FOR SEQ ID NO:  19:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2471 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  peptide
US-08-083-590A-19

  Query Match             2.8%;  Score 8;  DB 1;  Length 2471;
  Best Local Similarity   100.0%;  Pred. No. 62;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          5 PCQHGATC 12
              ||||||||
Db       1157 PCQHGATC 1164


RESULT 12
```

```
US-08-532-384-19
; Sequence 19, Application US/08532384
; Patent No. 6083904
;  GENERAL INFORMATION:
;    APPLICANT:  Artavanis-Tsakonas, S. et al.
;    TITLE OF INVENTION:  Therapeutic And Diagnostic Methods
;    TITLE OF INVENTION:  And Compositions Based On No. 6083904ch Proteins And
;    TITLE OF INVENTION:  Nucleic Acids
;    NUMBER OF SEQUENCES:  21
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Pennie & Edmonds
;      STREET:  1155 Avenue of the Americas
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  U.S.A.
;      ZIP:  10036
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/532,384
;      FILING DATE:
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/083,590
;      FILING DATE:  25-JUN-1993
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Misrock, S. Leslie
;      REGISTRATION NUMBER:  18,872
;      REFERENCE/DOCKET NUMBER:  7326-015
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  212 790-9090
;      TELEFAX:  212 8698864/9741
;      TELEX:  66141 PENNIE
;  INFORMATION FOR SEQ ID NO:  19:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2471 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  peptide
US-08-532-384-19

  Query Match             2.8%;  Score 8;  DB 3;  Length 2471;
  Best Local Similarity   100.0%;  Pred. No. 62;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          5 PCQHGATC 12
              ||||||||
Db       1157 PCQHGATC 1164


RESULT 13
US-08-899-232-1
; Sequence 1, Application US/08899232
; Patent No. 6436650
; GENERAL INFORMATION:
;  APPLICANT: Artavanis-Tsakonas, Spyridon
;  APPLICANT: Qi, Huilin
;  TITLE OF INVENTION: ACTIVATED FORMS OF NOTCH AND METHODS BASED THEREON
;  FILE REFERENCE: 7326-046
;  CURRENT APPLICATION NUMBER: US/08/899,232
;  CURRENT FILING DATE: 1997-07-23
;  NUMBER OF SEQ ID NOS: 4
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
;   LENGTH: 2471
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-08-899-232-1
```

```
  Query Match             2.8%;  Score 8;  DB 4;  Length 2471;
  Best Local Similarity   100.0%;  Pred. No. 62;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          5 PCQHGATC 12
              ||||||||
Db       1157 PCQHGATC 1164


RESULT 14
US-09-562-702A-12
; Sequence 12, Application US/09562702A
; Patent No. 6632790
; GENERAL INFORMATION:
;  APPLICANT: Yurchenco, Peter
;  TITLE OF INVENTION: Laminin 2 and Methods for Its Use
;  FILE REFERENCE: 99-274-B
;  CURRENT APPLICATION NUMBER: US/09/562,702A
;  CURRENT FILING DATE:  2000-04-28
;  PRIOR APPLICATION NUMBER: 60/155,945
;  PRIOR FILING DATE: 1999-09-24
;  PRIOR APPLICATION NUMBER: 60/143,289
;  PRIOR FILING DATE: 1999-07-12
;  PRIOR APPLICATION NUMBER: 60/139,198
;  PRIOR FILING DATE: 1999-06-15
;  PRIOR APPLICATION NUMBER: 60/131,720
;  PRIOR FILING DATE: 1999-04-30
;  NUMBER OF SEQ ID NOS: 32
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 12
;   LENGTH: 3084
;   TYPE: PRT
;   ORGANISM: Mus musculus
US-09-562-702A-12

  Query Match             2.8%;  Score 8;  DB 4;  Length 3084;
  Best Local Similarity   100.0%;  Pred. No. 76;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        122 TIELEVRT 129
              ||||||||
Db       2762 TIELEVRT 2769


RESULT 15
US-09-562-702A-8
; Sequence 8, Application US/09562702A
; Patent No. 6632790
; GENERAL INFORMATION:
;  APPLICANT: Yurchenco, Peter
;  TITLE OF INVENTION: Laminin 2 and Methods for Its Use
;  FILE REFERENCE: 99-274-B
;  CURRENT APPLICATION NUMBER: US/09/562,702A
;  CURRENT FILING DATE:  2000-04-28
;  PRIOR APPLICATION NUMBER: 60/155,945
;  PRIOR FILING DATE: 1999-09-24
;  PRIOR APPLICATION NUMBER: 60/143,289
;  PRIOR FILING DATE: 1999-07-12
;  PRIOR APPLICATION NUMBER: 60/139,198
;  PRIOR FILING DATE: 1999-06-15
;  PRIOR APPLICATION NUMBER: 60/131,720
;  PRIOR FILING DATE: 1999-04-30
;  NUMBER OF SEQ ID NOS: 32
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 8
;   LENGTH: 3088
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-562-702A-8

  Query Match             2.8%;  Score 8;  DB 4;  Length 3088;
  Best Local Similarity   100.0%;  Pred. No. 77;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        122 TIELEVRT 129
              ||||||||
Db       2766 TIELEVRT 2773

Search completed: March  9, 2004, 17:32:24
Job time : 10.8957 secs
```

```
                  GenCore version 5.1.6
          Copyright (c) 1993 - 2004  Compugen Ltd.

OM protein - protein search, using sw model

Run on:         March  9, 2004, 17:27:59 ; Search time 20.844 Seconds
                                            (without alignments)
                                            2876.963 Million cell updates/sec

Title:          US-10-006-011A-9
Perfect score:  284
Sequence:       1 CERQPCQHGATCMPAGEYEF..........QPLDLQHRAQAGANTRPCPS 284

Scoring table:  OLIGO
                Gapop 60.0 , Gapext 60.0

Searched:       809742 seqs, 211153259 residues

Word size :      0

Total number of hits satisfying chosen parameters:      809742

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :      Published_Applications_AA:*
             1:  /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep:*
             2:  /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep:*
             3:  /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep:*
             4:  /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep:*
             5:  /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep:*
             6:  /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep:*
             7:  /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep:*
             8:  /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep:*
             9:  /cgn2_6/ptodata/1/pubpaa/US09A_PUBCOMB.pep:*
            10:  /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pep:*
            11:  /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep:*
            12:  /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep:*
            13:  /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep:*
            14:  /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep:*
            15:  /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep:*
            16:  /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep:*
            17:  /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep:*
            18:  /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep:*

      Pred. No. is the number of results predicted by chance to have a
      score greater than or equal to the score of the result being printed,
      and is derived by analysis of the total score distribution.

                               SUMMARIES
                     %
Result             Query
  No.    Score    Match Length DB  ID                             Description
  ---------------------------------------------------------------------------
    1       9      3.2    1940 13  US-10-016-283-34               Sequence 34, Appl
    2       9      3.2    2531 15  US-10-190-115-29               Sequence 29, Appl
    3       9      3.2    2531 15  US-10-369-072-29               Sequence 29, Appl
    4       8      2.8      80  9  US-09-864-761-35784            Sequence 35784, A
    5       8      2.8    1307 14  US-10-303-685-17               Sequence 17, Appl
    6       8      2.8    1321 14  US-10-241-220-82               Sequence 82, Appl
    7       8      2.8    1321 15  US-10-295-027-262              Sequence 262, App
    8       8      2.8    1388 15  US-10-093-463-138              Sequence 138, App
    9       8      2.8    1388 15  US-10-093-463-140              Sequence 140, App
   10       8      2.8    1473 15  US-10-190-115-4                Sequence 4, Appli
   11       8      2.8    1473 15  US-10-369-072-4                Sequence 4, Appli
   12       8      2.8    2321 14  US-10-356-625-2                Sequence 2, Appli
   13       8      2.8    2447 15  US-10-190-115-28               Sequence 28, Appl
   14       8      2.8    2447 15  US-10-369-072-28               Sequence 28, Appl
   15       8      2.8    2469 15  US-10-190-115-2                Sequence 2, Appli
   16       8      2.8    2469 15  US-10-369-072-2                Sequence 2, Appli
   17       8      2.8    2471 15  US-10-190-115-27               Sequence 27, Appl
   18       8      2.8    2471 15  US-10-369-072-27               Sequence 27, Appl
   19       8      2.8    3070 10  US-09-961-403-7                Sequence 7, Appli
   20       7      2.5      12 14  US-10-084-994-18               Sequence 18, Appl
   21       7      2.5      12 14  US-10-193-109-18               Sequence 18, Appl
   22       7      2.5      12 15  US-10-193-409-18               Sequence 18, Appl
   23       7      2.5      29 14  US-10-029-386-32069            Sequence 32069, A
   24       7      2.5      43 14  US-10-050-704-269              Sequence 269, App
   25       7      2.5      50  9  US-09-975-143-1                Sequence 1, Appli
   26       7      2.5      61 14  US-10-074-475-177              Sequence 177, App
   27       7      2.5      72 14  US-10-106-698-6518             Sequence 6518, Ap
   28       7      2.5      76 11  US-09-864-408A-3218            Sequence 3218, Ap
   29       7      2.5      91 14  US-10-156-761-8320             Sequence 8320, Ap
   30       7      2.5     122 15  US-10-108-260A-3535            Sequence 3535, Ap
   31       7      2.5     152 15  US-10-104-047-3338             Sequence 3338, Ap
   32       7      2.5     157 14  US-10-156-761-11984            Sequence 11984, A
   33       7      2.5     193  9  US-09-809-545A-4               Sequence 4, Appli
   34       7      2.5     194  9  US-09-764-870-537              Sequence 537, App
   35       7      2.5     194 14  US-10-125-540-537              Sequence 537, App
   36       7      2.5     210  9  US-09-764-870-432              Sequence 432, App
   37       7      2.5     210 14  US-10-125-540-432              Sequence 432, App
   38       7      2.5     211 14  US-10-050-704-266              Sequence 266, App
   39       7      2.5     223 14  US-10-017-161-1538             Sequence 1538, Ap
   40       7      2.5     223 15  US-10-292-798-1232             Sequence 1232, Ap
   41       7      2.5     238  9  US-09-764-853-758              Sequence 758, App
   42       7      2.5     238  9  US-09-764-898-252              Sequence 252, App
   43       7      2.5     238 10  US-09-764-881-158              Sequence 158, App
   44       7      2.5     238 14  US-10-073-865-121              Sequence 121, App
   45       7      2.5     238 15  US-10-242-747-158              Sequence 158, App

                               ALIGNMENTS

RESULT 1
US-10-016-283-34
; Sequence 34, Application US/10016283
; Publication No. US20020164702A1
; GENERAL INFORMATION:
;  APPLICANT: Valenzuela et al., David M.
;  TITLE OF INVENTION: NOVEL TYROSINE KINASE RECEPTORS AND LIGANDS
;  FILE REFERENCE: REG195-B-PCT-US
;  CURRENT APPLICATION NUMBER: US/10/016,283
;  CURRENT FILING DATE:  2001-11-30
;  PRIOR APPLICATION NUMBER: US/09/077,955A
;  PRIOR FILING DATE: 1998-09-10
;  PRIOR APPLICATION NUMBER: PCT/US96/20696
;  PRIOR FILING DATE: 1996-12-13
;  NUMBER OF SEQ ID NOS: 36
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 34
;   LENGTH: 1940
;   TYPE: PRT
;   ORGANISM: Rattus sp.
US-10-016-283-34

  Query Match             3.2%;  Score 9;  DB 13;  Length 1940;
  Best Local Similarity   100.0%;  Pred. No. 16;
  Matches    9;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         45 PCLHGGTCQ 53
              |||||||||
Db       1228 PCLHGGTCQ 1236

RESULT 2
US-10-190-115-29
; Sequence 29, Application US/10190115
; Publication No. US20030207394A1
; GENERAL INFORMATION:
;  APPLICANT: Alsobrook, John P. II
;  APPLICANT:  Boldog, Ferenc L.
```

```
;  APPLICANT:  Burgess, Catherine E.
;  APPLICANT:  Casman, Stacie J.
;  APPLICANT:  Grosse, William M.
;  APPLICANT:  Gusev, Vladimir Y.
;  APPLICANT:  Ji, Weizhen
;  APPLICANT:  Lepley, Denise M.
;  APPLICANT:  Liu, Xiaohong
;  APPLICANT:  Mezick, Amanda J.
;  APPLICANT:  Padigaru, Muralidhara
;  APPLICANT:  Patturajan, Meera
;  APPLICANT:  Rastelli, Luca
;  APPLICANT:  Shen, Lei
;  APPLICANT:  Shenoy, Suresh G.
;  APPLICANT:  Shimkets, Richard A.
;  APPLICANT:  Spaderna, Steven K.
;  APPLICANT:  Spytek, Kimberly A.
;  APPLICANT:  Szekeres, Edward S. Jr.
;  APPLICANT:  Taupier, Raymond J. Jr.
;  APPLICANT:  Tchernev, Velizar T.
;  APPLICANT:  Zerhusen, Bryan D.
;  APPLICANT:  Voss, Edward Z.
;  TITLE OF INVENTION: NOVEL PROTEINS AND NUCLEIC ACIDS ENCODING SAME
;  FILE REFERENCE: 21402-050 CIP
;  CURRENT APPLICATION NUMBER: US/10/190,115
;  CURRENT FILING DATE:  2003-02-10
;  PRIOR APPLICATION NUMBER: 60/303,168
;  PRIOR FILING DATE: 2001-07-05
;  PRIOR APPLICATION NUMBER: 60/368,996
;  PRIOR FILING DATE: 2002-04-01
;  PRIOR APPLICATION NUMBER: 60/386,816
;  PRIOR FILING DATE: 2002-06-07
;  PRIOR APPLICATION NUMBER: 60/215,854
;  PRIOR FILING DATE: 2000-07-03
;  PRIOR APPLICATION NUMBER: 60/215,856
;  PRIOR FILING DATE: 2000-07-03
;  PRIOR APPLICATION NUMBER: 60/215,902
;  PRIOR FILING DATE: 2000-07-03
;  PRIOR APPLICATION NUMBER: 60/216,585,
;  PRIOR FILING DATE: 2000-07-07
;  PRIOR APPLICATION NUMBER: 60/216,586
;  PRIOR FILING DATE: 2001-07-07
;  PRIOR APPLICATION NUMBER: 60/216,722
;  PRIOR FILING DATE: 2000-07-07
;  PRIOR APPLICATION NUMBER: 60/218,622
;  PRIOR FILING DATE: 2000-07-17
;  Remaining Prior Application data removed - See File Wrapper or PALM.
;  NUMBER OF SEQ ID NOS: 136
;  SOFTWARE: CuraSeqList version 0.1
; SEQ ID NO 29
;   LENGTH: 2531
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-190-115-29

  Query Match              3.2%;  Score 9;  DB 15;  Length 2531;
  Best Local Similarity   100.0%;  Pred. No. 20;
  Matches    9;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         45 PCLHGGTCQ 53
              |||||||||
Db       1029 PCLHGGTCQ 1037

RESULT 3
US-10-369-072-29
; Sequence 29, Application US/10369072
; Publication No. US20040014081A1
; GENERAL INFORMATION:
;  APPLICANT: Alsobrook II, John P
;  APPLICANT:  Spaderna, Stephen K
;  APPLICANT:  Tchernev, Velizar
;  APPLICANT:  Liu, Xiaohong
;  APPLICANT:  Shenoy, Suresh
;  APPLICANT:  Spytek, Kimberly
;  APPLICANT:  Zerhusen, Bryan
;  APPLICANT:  Patturajan, Meera
;  APPLICANT:  Taupier, Raymond T
;  APPLICANT:  Rastelli, Luca
;  APPLICANT:  Grosse, William M
;  APPLICANT:  Szerkeres, Edward S
;  APPLICANT:  Lepley, Denise M
;  APPLICANT:  Shen, Lei
;  APPLICANT:  Burgess, Catherine E
;  APPLICANT:  Shimkets, Richard
;  APPLICANT:  Padigaru, Muralidhara
;  TITLE OF INVENTION: No. US20040014081A1el Proteins and Nucleic Acids Encoding Same
;  FILE REFERENCE: 21402-050 CON2
;  CURRENT APPLICATION NUMBER: US/10/369,072
;  CURRENT FILING DATE:  2003-02-18
;  PRIOR APPLICATION NUMBER: 10/174,372
;  PRIOR FILING DATE: 2002-06-17
;  PRIOR APPLICATION NUMBER: 09/898,994
;  PRIOR FILING DATE: 2001-07-03
;  PRIOR APPLICATION NUMBER: 60/215,854
;  PRIOR FILING DATE: 2000-07-03
;  PRIOR APPLICATION NUMBER: 60/215,856
;  PRIOR FILING DATE: 2000-07-03
;  PRIOR APPLICATION NUMBER: 60/215,902
;  PRIOR FILING DATE: 2000-07-03
;  PRIOR APPLICATION NUMBER: 60/216,585
;  PRIOR FILING DATE: 2000-07-07
;  PRIOR APPLICATION NUMBER: 60/216,586
;  PRIOR FILING DATE: 2000-07-07
;  PRIOR APPLICATION NUMBER: 60/216,722
;  PRIOR FILING DATE: 2000-07-07
;  PRIOR APPLICATION NUMBER: 60/218,622
;  PRIOR FILING DATE: 2000-07-17
;  PRIOR APPLICATION NUMBER: 60/218,992
;  PRIOR FILING DATE: 2000-07-17
;  Remaining Prior Application data removed - See File Wrapper or PALM.
;  NUMBER OF SEQ ID NOS: 100
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 29
;   LENGTH: 2531
;   TYPE: PRT
;   ORGANISM: Rattus norvegicus
US-10-369-072-29

  Query Match              3.2%;  Score 9;  DB 15;  Length 2531;
  Best Local Similarity   100.0%;  Pred. No. 20;
  Matches    9;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         45 PCLHGGTCQ 53
              |||||||||
Db       1029 PCLHGGTCQ 1037

RESULT 4
US-09-864-761-35784
; Sequence 35784, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
;  APPLICANT: Penn, Sharron G.
;  APPLICANT:  Rank, David R.
;  APPLICANT:  Hanzel, David K.
;  APPLICANT:  Chen, Wensheng
;  TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
;  TITLE OF INVENTION:  GENE EXPRESSION ANALYSIS BY MICROARRAY
;  FILE REFERENCE: Aeomica-X-1
;  CURRENT APPLICATION NUMBER: US/09/864,761
;  CURRENT FILING DATE:  2001-05-23
;  PRIOR APPLICATION NUMBER: US 60/180,312
;  PRIOR FILING DATE: 2000-02-04
;  PRIOR APPLICATION NUMBER: US 60/207,456
;  PRIOR FILING DATE: 2000-05-26
;  PRIOR APPLICATION NUMBER: US 09/632,366
```

```
;  PRIOR FILING DATE: 2000-08-03
;  PRIOR APPLICATION NUMBER: GB 24263.6
;  PRIOR FILING DATE: 2000-10-04
;  PRIOR APPLICATION NUMBER: US 60/236,359
;  PRIOR FILING DATE: 2000-09-27
;  PRIOR APPLICATION NUMBER: PCT/US01/00666
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00667
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00664
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00669
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00665
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00668
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00663
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00662
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00661
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00670
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: US 60/234,687
;  PRIOR FILING DATE: 2000-09-21
;  PRIOR APPLICATION NUMBER: US 09/608,408
;  PRIOR FILING DATE: 2000-06-30
;  PRIOR APPLICATION NUMBER: US 09/774,203
;  PRIOR FILING DATE: 2001-01-29
;  NUMBER OF SEQ ID NOS: 49117
;  SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 35784
;   LENGTH: 80
;   TYPE: PRT
;   ORGANISM: Homo sapiens
;   FEATURE:
;   OTHER INFORMATION: MAP TO AL035413.19
;   OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 4.1
;   OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 4.4
;   OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 3.2
;   OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 4.8
;   OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 6.5
;   OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL =  13
;   OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 5.8
;   OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 5.2
;   OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 6.2
;   OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL =   3
;   OTHER INFORMATION: EST_HUMAN HIT: BE875511.1, EVALUE 3.00e-14
US-09-864-761-35784

  Query Match             2.8%;  Score 8;  DB 9;  Length 80;
  Best Local Similarity   100.0%;  Pred. No. 8.9;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        259 PGAPPPQP 266
              ||||||||
Db         33 PGAPPPQP 40

RESULT 5
US-10-303-685-17
; Sequence 17, Application US/10303685
; Publication No. US20030100005A1
; GENERAL INFORMATION:
;  APPLICANT: Exelixis, Inc.
;  TITLE OF INVENTION: CRBs AS MODIFIERS OF BRANCHING MORPHOGENESIS AND METHODS OF USE
;  FILE REFERENCE: EX02-125C
;  CURRENT APPLICATION NUMBER: US/10/303,685
;  CURRENT FILING DATE:  2002-11-25
;  PRIOR APPLICATION NUMBER: 60/333,388
;  PRIOR FILING DATE: 2001-11-26
```

```
;  NUMBER OF SEQ ID NOS: 17
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 17
;   LENGTH: 1307
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-303-685-17

  Query Match             2.8%;  Score 8;  DB 14;  Length 1307;
  Best Local Similarity   100.0%;  Pred. No. 1e+02;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         45 PCLHGGTC 52
              ||||||||
Db        388 PCLHGGTC 395

RESULT 6
US-10-241-220-82
; Sequence 82, Application US/10241220
; Publication No. US20030148408A1
; GENERAL INFORMATION:
;  APPLICANT: Frantz,Gretchen
;  APPLICANT:  Hillan,Kenneth J.
;  APPLICANT:  Phillips,Heidi
;  APPLICANT:  Polakis,Paul
;  APPLICANT:  Spencer,Susan
;  APPLICANT:  Williams,P.Mickey
;  APPLICANT:  Wu,Thomas
;  APPLICANT:  Zhang,Zemin
;  TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE DIAGNOSIS AND
;  TITLE OF INVENTION:  TREATMENT OF TUMOR
;  FILE REFERENCE: P5010R1-US
;  CURRENT APPLICATION NUMBER: US/10/241,220
;  CURRENT FILING DATE:  2002-12-13
;  NUMBER OF SEQ ID NOS: 120
; SEQ ID NO 82
;   LENGTH: 1321
;   TYPE: PRT
;   ORGANISM: Homo Sapien
US-10-241-220-82

  Query Match             2.8%;  Score 8;  DB 14;  Length 1321;
  Best Local Similarity   100.0%;  Pred. No. 1e+02;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         45 PCLHGGTC 52
              ||||||||
Db       1016 PCLHGGTC 1023

RESULT 7
US-10-295-027-262
; Sequence 262, Application US/10295027
; Publication No. US20030232350A1
; GENERAL INFORMATION:
;  APPLICANT: Afar, Daniel
;  APPLICANT:  Aziz, Natasha
;  APPLICANT:  Ginsberg, Wendy M.
;  APPLICANT:  Gish, Kurt C.
;  APPLICANT:  Glynne, Richard
;  APPLICANT:  Hevezi, Peter A.
;  APPLICANT:  Mack, David H.
;  APPLICANT:  Murray, Richard
;  APPLICANT:  Watson, Susan R.
;  APPLICANT:  Eos Biotechnology, Inc.
;  TITLE OF INVENTION: Methods of Diagnosis of Cancer, Compositions and
;  TITLE OF INVENTION:  Methods of Screening for Modulators of Cancer
;  FILE REFERENCE: 018501-012500US
;  CURRENT APPLICATION NUMBER: US/10/295,027
;  CURRENT FILING DATE:  2002-11-13
;  PRIOR APPLICATION NUMBER: US 09/663,733
;  PRIOR FILING DATE: 2000-09-15
```

```
;  PRIOR APPLICATION NUMBER: US 60/350,666
;  PRIOR FILING DATE: 2001-11-13
;  PRIOR APPLICATION NUMBER: US 60/335,394
;  PRIOR FILING DATE: 2001-11-15
;  PRIOR APPLICATION NUMBER: US 60/332,464
;  PRIOR FILING DATE: 2001-11-21
;  PRIOR APPLICATION NUMBER: US 60/334,393
;  PRIOR FILING DATE: 2001-11-29
;  PRIOR APPLICATION NUMBER: US 60/340,376
;  PRIOR FILING DATE: 2001-12-14
;  PRIOR APPLICATION NUMBER: US 60/347,211
;  PRIOR FILING DATE: 2002-01-08
;  PRIOR APPLICATION NUMBER: US 60/347,349
;  PRIOR FILING DATE: 2002-01-10
;  PRIOR APPLICATION NUMBER: US 60/355,250
;  PRIOR FILING DATE: 2002-02-08
;  PRIOR APPLICATION NUMBER: US 60/356,714
;  PRIOR FILING DATE: 2002-02-13
;  Remaining Prior Application data removed - See File Wrapper or PALM.
;  NUMBER OF SEQ ID NOS: 1386
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 262
;   LENGTH: 1321
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-295-027-262

  Query Match             2.8%;  Score 8;  DB 15;  Length 1321;
  Best Local Similarity   100.0%;  Pred. No. 1e+02;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          45 PCLHGGTC 52
               ||||||||
Db        1016 PCLHGGTC 1023


RESULT 8
US-10-093-463-138
; Sequence 138, Application US/10093463
; Publication No. US20030208039A1
; GENERAL INFORMATION:
;  APPLICANT: Padigaru, Muralidhara
;  APPLICANT:  Shenoy, Suresh
;  APPLICANT:  Kekuda, Ramesh
;  APPLICANT:  Gusev, Vladimir
;  APPLICANT:  Pochart, Pascal
;  APPLICANT:  Zhong, Mei
;  APPLICANT:  Rastelli, Luca
;  APPLICANT:  Mezes, Peter
;  APPLICANT:  Smithson, Glennda
;  APPLICANT:  Guo, Xiaojia
;  APPLICANT:  Gerlach, Valerie
;  APPLICANT:  Casman, Stacie
;  APPLICANT:  Boldog, Ferenc
;  APPLICANT:  Li, Li
;  APPLICANT:  Zerhusen, Bryan
;  APPLICANT:  Tchernev, Velizar
;  APPLICANT:  Gangolli, Esha
;  APPLICANT:  Vernet, Corine
;  APPLICANT:  Pena, Carol
;  APPLICANT:  Burgess, Catherine
;  APPLICANT:  Liu, Xiaohong
;  APPLICANT:  Spytek, Kimberly
;  APPLICANT:  Gorman, Linda
;  APPLICANT:  Spaderna, Steven
;  APPLICANT:  Voss, Edward
;  APPLICANT:  Malyankar, Uriel
;  APPLICANT:  Anderson, David
;  APPLICANT:  Patturajan, Meera
;  APPLICANT:  Miller, Charles
;  APPLICANT:  Taupier, Raymond J. Jr.
;  TITLE OF INVENTION: No. US20030208039A1el Antibodies that Bind to Antigenic Polypepti
;  TITLE OF INVENTION:  Encoding The Antigens, and Methods of Use.
```

```
;  FILE REFERENCE: 21402-290A (Cura 590AT)
;  CURRENT APPLICATION NUMBER: US/10/093,463
;  CURRENT FILING DATE:  2002-06-24
;  PRIOR APPLICATION NUMBER: 60/283,675
;  PRIOR FILING DATE: 2001-04-14
;  PRIOR APPLICATION NUMBER: 60/338,092
;  PRIOR FILING DATE: 2001-12-03
;  PRIOR APPLICATION NUMBER: 60/274,281
;  PRIOR FILING DATE: 2001-03-08
;  PRIOR APPLICATION NUMBER: 60/274,101
;  PRIOR FILING DATE: 2001-03-08
;  PRIOR APPLICATION NUMBER: 60/325,681
;  PRIOR FILING DATE: 2001-09-27
;  PRIOR APPLICATION NUMBER: 60/304,354
;  PRIOR FILING DATE: 2001-07-10
;  PRIOR APPLICATION NUMBER: 60/279,995
;  PRIOR FILING DATE: 2001-03-30
;  PRIOR APPLICATION NUMBER: 60/294,899
;  PRIOR FILING DATE: 2001-05-31
;  PRIOR APPLICATION NUMBER: 60/287,424
;  PRIOR FILING DATE: 2001-04-30
;  PRIOR APPLICATION NUMBER: 60/299,027
;  PRIOR FILING DATE: 2001-06-18
;  PRIOR APPLICATION NUMBER: 60/309,198
;  PRIOR FILING DATE: 2001-07-31
;  PRIOR APPLICATION NUMBER: 60/281,194
;  PRIOR FILING DATE: 2001-04-04
;  PRIOR APPLICATION NUMBER: 60/274,194
;  PRIOR FILING DATE: 2001-03-08
;  PRIOR APPLICATION NUMBER: 60/274,849
;  PRIOR FILING DATE: 2001-03-09
;  PRIOR APPLICATION NUMBER: 60/330,380
;  PRIOR FILING DATE: 2001-10-18
;  PRIOR APPLICATION NUMBER: 60/275,235
;  PRIOR FILING DATE: 2001-03-12
;  PRIOR APPLICATION NUMBER: 60/288,342
;  PRIOR FILING DATE: 2001-05-03
;  PRIOR APPLICATION NUMBER: 60/275,578
;  PRIOR FILING DATE: 2001-03-13
;  NUMBER OF SEQ ID NOS: 370
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 138
;   LENGTH: 1388
;   TYPE: PRT
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: Variant
;   LOCATION: (113)...(113)
;   OTHER INFORMATION: Wherein Xaa is any amino acid as defined in the specification.
US-10-093-463-138

  Query Match             2.8%;  Score 8;  DB 15;  Length 1388;
  Best Local Similarity   100.0%;  Pred. No. 1.1e+02;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          45 PCLHGGTC 52
               ||||||||
Db         469 PCLHGGTC 476


RESULT 9
US-10-093-463-140
; Sequence 140, Application US/10093463
; Publication No. US20030208039A1
; GENERAL INFORMATION:
;  APPLICANT: Padigaru, Muralidhara
;  APPLICANT:  Shenoy, Suresh
;  APPLICANT:  Kekuda, Ramesh
;  APPLICANT:  Gusev, Vladimir
;  APPLICANT:  Pochart, Pascal
;  APPLICANT:  Zhong, Mei
;  APPLICANT:  Rastelli, Luca
;  APPLICANT:  Mezes, Peter
```

```
;  APPLICANT:  Smithson, Glennda
;  APPLICANT:  Guo, Xiaojia
;  APPLICANT:  Gerlach, Valerie
;  APPLICANT:  Casman, Stacie
;  APPLICANT:  Boldog, Ferenc
;  APPLICANT:  Li, Li
;  APPLICANT:  Zerhusen, Bryan
;  APPLICANT:  Tchernev, Velizar
;  APPLICANT:  Gangolli, Esha
;  APPLICANT:  Vernet, Corine
;  APPLICANT:  Pena, Carol
;  APPLICANT:  Burgess, Catherine
;  APPLICANT:  Liu, Xiaohong
;  APPLICANT:  Spytek, Kimberly
;  APPLICANT:  Gorman, Linda
;  APPLICANT:  Spaderna, Steven
;  APPLICANT:  Voss, Edward
;  APPLICANT:  Malyankar, Uriel
;  APPLICANT:  Anderson, David
;  APPLICANT:  Patturajan, Meera
;  APPLICANT:  Miller, Charles
;  APPLICANT:  Taupier, Raymond J. Jr.
;  TITLE OF INVENTION: No. US20030208039A1el Antibodies that Bind to Antigenic Polypepti
;  TITLE OF INVENTION:  Encoding The Antigens, and Methods of Use.
;  FILE REFERENCE: 21402-290A (Cura 590AT)
;  CURRENT APPLICATION NUMBER: US/10/093,463
;  CURRENT FILING DATE:  2002-06-24
;  PRIOR APPLICATION NUMBER: 60/283,675
;  PRIOR FILING DATE: 2001-04-14
;  PRIOR APPLICATION NUMBER: 60/338,092
;  PRIOR FILING DATE: 2001-12-03
;  PRIOR APPLICATION NUMBER: 60/274,281
;  PRIOR FILING DATE: 2001-03-08
;  PRIOR APPLICATION NUMBER: 60/274,101
;  PRIOR FILING DATE: 2001-03-08
;  PRIOR APPLICATION NUMBER: 60/325,681
;  PRIOR FILING DATE: 2001-09-27
;  PRIOR APPLICATION NUMBER: 60/304,354
;  PRIOR FILING DATE: 2001-07-10
;  PRIOR APPLICATION NUMBER: 60/279,995
;  PRIOR FILING DATE: 2001-03-30
;  PRIOR APPLICATION NUMBER: 60/294,899
;  PRIOR FILING DATE: 2001-05-31
;  PRIOR APPLICATION NUMBER: 60/287,424
;  PRIOR FILING DATE: 2001-04-30
;  PRIOR APPLICATION NUMBER: 60/299,027
;  PRIOR FILING DATE: 2001-06-18
;  PRIOR APPLICATION NUMBER: 60/309,198
;  PRIOR FILING DATE: 2001-07-31
;  PRIOR APPLICATION NUMBER: 60/281,194
;  PRIOR FILING DATE: 2001-04-04
;  PRIOR APPLICATION NUMBER: 60/274,194
;  PRIOR FILING DATE: 2001-03-08
;  PRIOR APPLICATION NUMBER: 60/274,849
;  PRIOR FILING DATE: 2001-03-09
;  PRIOR APPLICATION NUMBER: 60/330,380
;  PRIOR FILING DATE: 2001-10-18
;  PRIOR APPLICATION NUMBER: 60/275,235
;  PRIOR FILING DATE: 2001-03-12
;  PRIOR APPLICATION NUMBER: 60/288,342
;  PRIOR FILING DATE: 2001-05-03
;  PRIOR APPLICATION NUMBER: 60/275,578
;  PRIOR FILING DATE: 2001-03-13
;  NUMBER OF SEQ ID NOS: 370
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 140
;   LENGTH: 1388
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-093-463-140

  Query Match             2.8%;  Score 8;  DB 15;  Length 1388;
  Best Local Similarity   100.0%;  Pred. No. 1.1e+02;
```

```
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         45 PCLHGGTC 52
              ||||||||
Db        469 PCLHGGTC 476


RESULT 10
US-10-190-115-4
; Sequence 4, Application US/10190115
; Publication No. US20030207394A1
; GENERAL INFORMATION:
;  APPLICANT: Alsobrook, John P. II
;  APPLICANT:  Boldog, Ferenc L.
;  APPLICANT:  Burgess, Catherine E.
;  APPLICANT:  Casman, Stacie J.
;  APPLICANT:  Grosse, William M.
;  APPLICANT:  Gusev, Vladimir Y.
;  APPLICANT:  Ji, Weizhen
;  APPLICANT:  Lepley, Denise M.
;  APPLICANT:  Liu, Xiaohong
;  APPLICANT:  Mezick, Amanda J.
;  APPLICANT:  Padigaru, Muralidhara
;  APPLICANT:  Patturajan, Meera
;  APPLICANT:  Rastelli, Luca
;  APPLICANT:  Shen, Lei
;  APPLICANT:  Shenoy, Suresh G.
;  APPLICANT:  Shimkets, Richard A.
;  APPLICANT:  Spaderna, Steven K.
;  APPLICANT:  Spytek, Kimberly A.
;  APPLICANT:  Szekeres, Edward S. Jr.
;  APPLICANT:  Taupier, Raymond J. Jr.
;  APPLICANT:  Tchernev, Velizar T.
;  APPLICANT:  Zerhusen, Bryan D.
;  APPLICANT:  Voss, Edward Z.
;  TITLE OF INVENTION: NOVEL PROTEINS AND NUCLEIC ACIDS ENCODING SAME
;  FILE REFERENCE: 21402-050 CIP
;  CURRENT APPLICATION NUMBER: US/10/190,115
;  CURRENT FILING DATE:  2003-02-10
;  PRIOR APPLICATION NUMBER: 60/303,168
;  PRIOR FILING DATE: 2001-07-05
;  PRIOR APPLICATION NUMBER: 60/368,996
;  PRIOR FILING DATE: 2002-04-01
;  PRIOR APPLICATION NUMBER: 60/386,816
;  PRIOR FILING DATE: 2002-06-07
;  PRIOR APPLICATION NUMBER: 60/215,854
;  PRIOR FILING DATE: 2000-07-03
;  PRIOR APPLICATION NUMBER: 60/215,856
;  PRIOR FILING DATE: 2000-07-03
;  PRIOR APPLICATION NUMBER: 60/215,902
;  PRIOR FILING DATE: 2000-07-03
;  PRIOR APPLICATION NUMBER: 60/216,585,
;  PRIOR FILING DATE: 2000-07-07
;  PRIOR APPLICATION NUMBER: 60/216,586
;  PRIOR FILING DATE: 2001-07-07
;  PRIOR APPLICATION NUMBER: 60/216,722
;  PRIOR FILING DATE: 2000-07-07
;  PRIOR APPLICATION NUMBER: 60/218,622
;  PRIOR FILING DATE: 2000-07-17
;  Remaining Prior Application data removed - See File Wrapper or PALM.
;  NUMBER OF SEQ ID NOS: 136
;  SOFTWARE: CuraSeqList version 0.1
; SEQ ID NO 4
;   LENGTH: 1473
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-190-115-4

  Query Match             2.8%;  Score 8;  DB 15;  Length 1473;
  Best Local Similarity   100.0%;  Pred. No. 1.1e+02;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          5 PCQHGATC 12
```

```
              ||||||||
Db       1157 PCQHGATC 1164


RESULT 11
US-10-369-072-4
; Sequence 4, Application US/10369072
; Publication No. US20040014081A1
; GENERAL INFORMATION:
;  APPLICANT: Alsobrook II, John P
;  APPLICANT:  Spaderna, Stephen K
;  APPLICANT:  Tchernev, Velizar
;  APPLICANT:  Liu, Xiaohong
;  APPLICANT:  Shenoy, Suresh
;  APPLICANT:  Spytek, Kimberly
;  APPLICANT:  Zerhusen, Bryan
;  APPLICANT:  Patturajan, Meera
;  APPLICANT:  Taupier, Raymond T
;  APPLICANT:  Rastelli, Luca
;  APPLICANT:  Grosse, William M
;  APPLICANT:  Szerkeres, Edward S
;  APPLICANT:  Lepley, Denise M
;  APPLICANT:  Shen, Lei
;  APPLICANT:  Burgess, Catherine E
;  APPLICANT:  Shimkets, Richard
;  APPLICANT:  Padigaru, Muralidhara
;  TITLE OF INVENTION: No. US20040014081A1el Proteins and Nucleic Acids Encoding Same
;  FILE REFERENCE: 21402-050 CON2
;  CURRENT APPLICATION NUMBER: US/10/369,072
;  CURRENT FILING DATE:  2003-02-18
;  PRIOR APPLICATION NUMBER: 10/174,372
;  PRIOR FILING DATE: 2002-06-17
;  PRIOR APPLICATION NUMBER: 09/898,994
;  PRIOR FILING DATE: 2001-07-03
;  PRIOR APPLICATION NUMBER: 60/215,854
;  PRIOR FILING DATE: 2000-07-03
;  PRIOR APPLICATION NUMBER: 60/215,856
;  PRIOR FILING DATE: 2000-07-03
;  PRIOR APPLICATION NUMBER: 60/215,902
;  PRIOR FILING DATE: 2000-07-03
;  PRIOR APPLICATION NUMBER: 60/216,585
;  PRIOR FILING DATE: 2000-07-07
;  PRIOR APPLICATION NUMBER: 60/216,586
;  PRIOR FILING DATE: 2000-07-07
;  PRIOR APPLICATION NUMBER: 60/216,722
;  PRIOR FILING DATE: 2000-07-07
;  PRIOR APPLICATION NUMBER: 60/218,622
;  PRIOR FILING DATE: 2000-07-17
;  PRIOR APPLICATION NUMBER: 60/218,992
;  PRIOR FILING DATE: 2000-07-17
;  Remaining Prior Application data removed - See File Wrapper or PALM.
;  NUMBER OF SEQ ID NOS: 100
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
;   LENGTH: 1473
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-369-072-4

  Query Match             2.8%;  Score 8;  DB 15;  Length 1473;
  Best Local Similarity   100.0%;  Pred. No. 1.1e+02;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          5 PCQHGATC 12
              ||||||||
Db       1157 PCQHGATC 1164


RESULT 12
US-10-356-625-2
; Sequence 2, Application US/10356625
; Publication No. US20030186290A1
; GENERAL INFORMATION:
```

```
;  APPLICANT: Tournier-Lasserve, Elisabeth
;  APPLICANT:  Joutel, Anne
;  APPLICANT:  Bousser, Marie-Germaine
;  APPLICANT:  Bach, Jean-Francois
;  TITLE OF INVENTION: GENE INVOLVED IN CADASIL, METHOD OF DIAGNOSIS AND
;  TITLE OF INVENTION:  THERAPEUTIC APPLICATION
;  FILE REFERENCE: 03715.0048-00000
;  CURRENT APPLICATION NUMBER: US/10/356,625
;  CURRENT FILING DATE:  2003-02-03
;  PRIOR APPLICATION NUMBER: US/09/230,652
;  PRIOR FILING DATE: 1999-05-17
;  PRIOR APPLICATION NUMBER: FR 96 09733
;  PRIOR FILING DATE: 1996-08-01
;  PRIOR APPLICATION NUMBER: FR 97 04680
;  PRIOR FILING DATE: 1997-04-16
;  PRIOR APPLICATION NUMBER: PCT/FR97/01433
;  PRIOR FILING DATE: 1997-07-31
;  NUMBER OF SEQ ID NOS: 163
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
;   LENGTH: 2321
;   TYPE: PRT
;   ORGANISM: Homo sapiens
;   FEATURE:
;   OTHER INFORMATION: human ADNc No. US20030186290A1ch 3
US-10-356-625-2

  Query Match             2.8%;  Score 8;  DB 14;  Length 2321;
  Best Local Similarity   100.0%;  Pred. No. 1.7e+02;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         62 GFSGPRCQ 69
              ||||||||
Db       1237 GFSGPRCQ 1244


RESULT 13
US-10-190-115-28
; Sequence 28, Application US/10190115
; Publication No. US20030207394A1
; GENERAL INFORMATION:
;  APPLICANT: Alsobrook, John P. II
;  APPLICANT:  Boldog, Ferenc L.
;  APPLICANT:  Burgess, Catherine E.
;  APPLICANT:  Casman, Stacie J.
;  APPLICANT:  Grosse, William M.
;  APPLICANT:  Gusev, Vladimir Y.
;  APPLICANT:  Ji, Weizhen
;  APPLICANT:  Lepley, Denise M.
;  APPLICANT:  Liu, Xiaohong
;  APPLICANT:  Mezick, Amanda J.
;  APPLICANT:  Padigaru, Muralidhara
;  APPLICANT:  Patturajan, Meera
;  APPLICANT:  Rastelli, Luca
;  APPLICANT:  Shen, Lei
;  APPLICANT:  Shenoy, Suresh G.
;  APPLICANT:  Shimkets, Richard A.
;  APPLICANT:  Spaderna, Steven K.
;  APPLICANT:  Spytek, Kimberly A.
;  APPLICANT:  Szekeres, Edward S. Jr.
;  APPLICANT:  Taupier, Raymond J. Jr.
;  APPLICANT:  Tchernev, Velizar T.
;  APPLICANT:  Zerhusen, Bryan D.
;  APPLICANT:  Voss, Edward Z.
;  TITLE OF INVENTION: NOVEL PROTEINS AND NUCLEIC ACIDS ENCODING SAME
;  FILE REFERENCE: 21402-050 CIP
;  CURRENT APPLICATION NUMBER: US/10/190,115
;  CURRENT FILING DATE:  2003-02-10
;  PRIOR APPLICATION NUMBER: 60/303,168
;  PRIOR FILING DATE: 2001-07-05
;  PRIOR APPLICATION NUMBER: 60/368,996
;  PRIOR FILING DATE: 2002-04-01
;  PRIOR APPLICATION NUMBER: 60/386,816
```

```
;  PRIOR FILING DATE: 2002-06-07
;  PRIOR APPLICATION NUMBER: 60/215,854
;  PRIOR FILING DATE: 2000-07-03
;  PRIOR APPLICATION NUMBER: 60/215,856
;  PRIOR FILING DATE: 2000-07-03
;  PRIOR APPLICATION NUMBER: 60/215,902
;  PRIOR FILING DATE: 2000-07-03
;  PRIOR APPLICATION NUMBER: 60/216,585,
;  PRIOR FILING DATE: 2000-07-07
;  PRIOR APPLICATION NUMBER: 60/216,586
;  PRIOR FILING DATE: 2001-07-07
;  PRIOR APPLICATION NUMBER: 60/216,722
;  PRIOR FILING DATE: 2000-07-07
;  PRIOR APPLICATION NUMBER: 60/218,622
;  PRIOR FILING DATE: 2000-07-17
;  Remaining Prior Application data removed - See File Wrapper or PALM.
;  NUMBER OF SEQ ID NOS: 136
;  SOFTWARE: CuraSeqList version 0.1
; SEQ ID NO 28
;   LENGTH: 2447
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-190-115-28

  Query Match             2.8%;  Score 8;  DB 15;  Length 2447;
  Best Local Similarity   100.0%;  Pred. No. 1.8e+02;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         45 PCLHGGTC 52
              ||||||||
Db       1150 PCLHGGTC 1157


RESULT 14
US-10-369-072-28
; Sequence 28, Application US/10369072
; Publication No. US20040014081A1
; GENERAL INFORMATION:
;  APPLICANT: Alsobrook II, John P
;  APPLICANT:  Spaderna, Stephen K
;  APPLICANT:  Tchernev, Velizar
;  APPLICANT:  Liu, Xiaohong
;  APPLICANT:  Shenoy, Suresh
;  APPLICANT:  Spytek, Kimberly
;  APPLICANT:  Zerhusen, Bryan
;  APPLICANT:  Patturajan, Meera
;  APPLICANT:  Taupier, Raymond T
;  APPLICANT:  Rastelli, Luca
;  APPLICANT:  Grosse, William M
;  APPLICANT:  Szerkeres, Edward S
;  APPLICANT:  Lepley, Denise M
;  APPLICANT:  Shen, Lei
;  APPLICANT:  Burgess, Catherine E
;  APPLICANT:  Shimkets, Richard
;  APPLICANT:  Padigaru, Muralidhara
;  TITLE OF INVENTION: No. US20040014081A1el Proteins and Nucleic Acids Encoding Same
;  FILE REFERENCE: 21402-050 CON2
;  CURRENT APPLICATION NUMBER: US/10/369,072
;  CURRENT FILING DATE:  2003-02-18
;  PRIOR APPLICATION NUMBER: 10/174,372
;  PRIOR FILING DATE: 2002-06-17
;  PRIOR APPLICATION NUMBER: 09/898,994
;  PRIOR FILING DATE: 2001-07-03
;  PRIOR APPLICATION NUMBER: 60/215,854
;  PRIOR FILING DATE: 2000-07-03
;  PRIOR APPLICATION NUMBER: 60/215,856
;  PRIOR FILING DATE: 2000-07-03
;  PRIOR APPLICATION NUMBER: 60/215,902
;  PRIOR FILING DATE: 2000-07-03
;  PRIOR APPLICATION NUMBER: 60/216,585
;  PRIOR FILING DATE: 2000-07-07
;  PRIOR APPLICATION NUMBER: 60/216,586
;  PRIOR FILING DATE: 2000-07-07
```

```
;  PRIOR APPLICATION NUMBER: 60/216,722
;  PRIOR FILING DATE: 2000-07-07
;  PRIOR APPLICATION NUMBER: 60/218,622
;  PRIOR FILING DATE: 2000-07-17
;  PRIOR APPLICATION NUMBER: 60/218,992
;  PRIOR FILING DATE: 2000-07-17
;  Remaining Prior Application data removed - See File Wrapper or PALM.
;  NUMBER OF SEQ ID NOS: 100
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 28
;   LENGTH: 2447
;   TYPE: PRT
;   ORGANISM: Takifugu rubripes
US-10-369-072-28

  Query Match             2.8%;  Score 8;  DB 15;  Length 2447;
  Best Local Similarity   100.0%;  Pred. No. 1.8e+02;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         45 PCLHGGTC 52
              ||||||||
Db       1150 PCLHGGTC 1157


RESULT 15
US-10-190-115-2
; Sequence 2, Application US/10190115
; Publication No. US20030207394A1
; GENERAL INFORMATION:
;  APPLICANT: Alsobrook, John P. II
;  APPLICANT:  Boldog, Ferenc L.
;  APPLICANT:  Burgess, Catherine E.
;  APPLICANT:  Casman, Stacie J.
;  APPLICANT:  Grosse, William M.
;  APPLICANT:  Gusev, Vladimir Y.
;  APPLICANT:  Ji, Weizhen
;  APPLICANT:  Lepley, Denise M.
;  APPLICANT:  Liu, Xiaohong
;  APPLICANT:  Mezick, Amanda J.
;  APPLICANT:  Padigaru, Muralidhara
;  APPLICANT:  Patturajan, Meera
;  APPLICANT:  Rastelli, Luca
;  APPLICANT:  Shen, Lei
;  APPLICANT:  Shenoy, Suresh G.
;  APPLICANT:  Shimkets, Richard A.
;  APPLICANT:  Spaderna, Steven K.
;  APPLICANT:  Spytek, Kimberly A.
;  APPLICANT:  Szekeres, Edward S. Jr.
;  APPLICANT:  Taupier, Raymond J. Jr.
;  APPLICANT:  Tchernev, Velizar T.
;  APPLICANT:  Zerhusen, Bryan D.
;  APPLICANT:  Voss, Edward Z.
;  TITLE OF INVENTION: NOVEL PROTEINS AND NUCLEIC ACIDS ENCODING SAME
;  FILE REFERENCE: 21402-050 CIP
;  CURRENT APPLICATION NUMBER: US/10/190,115
;  CURRENT FILING DATE:  2003-02-10
;  PRIOR APPLICATION NUMBER: 60/303,168
;  PRIOR FILING DATE: 2001-07-05
;  PRIOR APPLICATION NUMBER: 60/368,996
;  PRIOR FILING DATE: 2002-04-01
;  PRIOR APPLICATION NUMBER: 60/386,816
;  PRIOR FILING DATE: 2002-06-07
;  PRIOR APPLICATION NUMBER: 60/215,854
;  PRIOR FILING DATE: 2000-07-03
;  PRIOR APPLICATION NUMBER: 60/215,856
;  PRIOR FILING DATE: 2000-07-03
;  PRIOR APPLICATION NUMBER: 60/215,902
;  PRIOR FILING DATE: 2000-07-03
;  PRIOR APPLICATION NUMBER: 60/216,585,
;  PRIOR FILING DATE: 2000-07-07
;  PRIOR APPLICATION NUMBER: 60/216,586
;  PRIOR FILING DATE: 2001-07-07
;  PRIOR APPLICATION NUMBER: 60/216,722
```

```
;  PRIOR FILING DATE: 2000-07-07
;  PRIOR APPLICATION NUMBER: 60/218,622
;  PRIOR FILING DATE: 2000-07-17
;  Remaining Prior Application data removed - See File Wrapper or PALM.
;  NUMBER OF SEQ ID NOS: 136
;  SOFTWARE: CuraSeqList version 0.1
; SEQ ID NO 2
;   LENGTH: 2469
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-190-115-2

  Query Match             2.8%;  Score 8;  DB 15;  Length 2469;
  Best Local Similarity   100.0%;  Pred. No. 1.8e+02;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          5 PCQHGATC 12
              ||||||||
Db       1156 PCQHGATC 1163


Search completed: March  9, 2004, 17:34:06
Job time : 21.844 secs
```

```
                    GenCore version 5.1.6
          Copyright (c) 1993 - 2004  Compugen Ltd.

OM protein - protein search, using sw model

Run on:         March  9, 2004, 17:23:39 ; Search time 9.71143 Seconds
                                           (without alignments)
                                           2813.016 Million cell updates/sec

Title:          US-10-006-011A-9
Perfect score:  284
Sequence:       1 CERQPCQHGATCMPAGEYEF..........QPLDLQHRAQAGANTRPCPS 284

Scoring table:  OLIGO
                Gapop 60.0 , Gapext 60.0

Searched:       283366 seqs, 96191526 residues

Word size :       0

Total number of hits satisfying chosen parameters:      283366

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :        PIR_78:*
                1:  pir1:*
                2:  pir2:*
                3:  pir3:*
                4:  pir4:*

      Pred. No. is the number of results predicted by chance to have a
      score greater than or equal to the score of the result being printed,
      and is derived by analysis of the total score distribution.

                                    SUMMARIES
                    %
Result              Query
  No.     Score     Match Length DB  ID                                   Description
 -------------------------------------------------------------------------------------
     1      284    100.0    4391  2  A38096                               perlecan precursor
     2       28      9.9    3707  2  S18252                               heparan sulfate pr
     3        9      3.2     861  2  A48825                               Notch homolog Motc
     4        9      3.2    1959  1  AGRT                                 agrin - rat
     5        9      3.2    2531  2  S18188                               notch protein homo
     6        9      3.2    2531  2  A46019                               notch-1 protein -
     7        8      2.8     497  2  AH2015                               sodium/solute symp
     8        8      2.8     500  2  F70012                               leucyl aminopeptid
     9        8      2.8    1203  2  A49175                               Motch B protein -
    10        8      2.8    1257  2  S28764                               neurocan precursor
    11        8      2.8    1268  2  S52781                               neurocan - mouse
    12        8      2.8    1328  2  T43060                               agrin - electric r
    13        8      2.8    1751  1  MMHUMH                               laminin alpha-2 ch
    14        8      2.8    1955  1  AGCH                                 agrin precursor -
    15        8      2.8    2321  2  S78549                               notch3 protein - h
    16        8      2.8    2471  2  A49128                               cell-fate determin
    17        8      2.8    3106  1  S53868                               laminin alpha-2 ch
    18        7      2.5      93  2  S75571                               hypothetical prote
    19        7      2.5     181  2  E87418                               chemotaxis protein
    20        7      2.5     201  2  E75567                               hypothetical prote
    21        7      2.5     210  2  T03144                               hypothetical prote
    22        7      2.5     221  2  T24494                               hypothetical prote
    23        7      2.5     224  2  T10660                               photosystem II pro
    24        7      2.5     239  2  AE0347                               two-component syst
    25        7      2.5     240  1  JS0591                               endo-1,4-beta-xyla
    26        7      2.5     240  1  S47512                               endo-1,4-beta-xyla
    27        7      2.5     240  2  A82783                               hypothetical prote
    28        7      2.5     241  2  T37005                               endo-1,4-beta-xyla
    29        7      2.5     244  2  T24631                               hypothetical prote
    30        7      2.5     255  2  A35255                               chlorocatechol 1,2
    31        7      2.5     257  2  AB2771                               hypothetical prote
    32        7      2.5     257  2  B97551                               general 1-amino ac
    33        7      2.5     273  2  B71355                               probable ribosomal
    34        7      2.5     275  2  AH0335                               probable penicilli
    35        7      2.5     283  2  F83874                               nickel ABC transpo
    36        7      2.5     298  2  D83033                               probable transcrip
    37        7      2.5     305  2  B55228                               methionyl-tRNA for
    38        7      2.5     330  2  T46256                               brevican - human (
    39        7      2.5     347  2  T35013                               probable membrane
    40        7      2.5     348  2  B58892                               NADH2 dehydrogenas
    41        7      2.5     373  2  F72602                               probable acetylpol
    42        7      2.5     399  2  T01387                               oxidoreductase hom
    43        7      2.5     400  2  D75331                               hypothetical prote
    44        7      2.5     401  2  S65138                               glycoprotein antig
    45        7      2.5     413  2  T29505                               hypothetical prote

                                 ALIGNMENTS

RESULT 1
A38096
perlecan precursor - human
N;Alternate names: basement membrane heparan sulfate proteoglycan; heparan sulfate prote
C;Species: Homo sapiens (man)
C;Date: 07-Apr-1994 #sequence_revision 07-Apr-1994 #text_change 05-Nov-1999
C;Accession: A38096; S19256; S77946; A41059; A40306; B33625; A33625; A41736
R;Murdoch, A.D.; Dodge, G.R.; Cohen, I.; Tuan, R.S.; Iozzo, R.V.
J. Biol. Chem. 267, 8544-8557, 1992
A;Title: Primary structure of the human heparan sulfate proteoglycan from basement membr
tor, laminin, neural cell adhesion molecules, and epidermal growth factor.
A;Reference number: A38096; MUID:92235084; PMID:1569102
A;Accession: A38096
A;Molecule type: mRNA
A;Residues: 1-4391 <MUR>
A;Cross-references: GB:M85289; NID:g184426; PIDN:AAA52700.1; PID:g184427
R;Kallunki, P.; Tryggvason, K.
J. Cell Biol. 116, 559-571, 1992
A;Title: Human basement membrane heparan sulfate proteoglycan core protein: a 467-kD pro
ell adhesion molecules, and epidermal growth factor.
A;Reference number: A41736; MUID:92112994; PMID:1730768
A;Accession: S19256
A;Molecule type: mRNA
A;Residues: 1-57,'D',59-434,'A',436,'FL',438-449,'Q',451-502,'A',503-792,'K',794-908,'R'
71-2979,'H',2981-2994,'G',2996-3167,'T',3169-3240,'R',3242-3426,'R',3428-3631,'Q',3633-3
A;Cross-references: EMBL:X62515
R;Tryggvason, K.
submitted to the EMBL Data Library, October 1991
A;Reference number: S77946
A;Accession: S77946
A;Molecule type: mRNA
A;Residues: 1-57,'D',59-434,'A',436,'FL',438-449,'Q',451-502,'A',503-792,'K',794-908,'R'
71-2979,'H',2981-2994,'G',2996-3167,'T',3169-3240,'R',3242-3426,'R',3428-3631,'Q',3633-4
A;Cross-references: EMBL:X62515; NID:g29469; PIDN:CAA44373.1; PID:g29470
R;Kallunki, P.; Eddy, R.L.; Byers, M.G.; Kestilae, M.; Shows, T.B.; Tryggvason, K.
Genomics 11, 389-396, 1991
A;Title: Cloning of human heparan sulfate proteoglycan core protein, assignment of the g
A;Reference number: A41059; MUID:92120660; PMID:1685141
A;Accession: A41059
A;Molecule type: mRNA
A;Residues: 'RT',892-908,'R',910-1101,'L',1103-1132,'L',1134-1221,'L',1223-1397 <KA2>
A;Cross-references: GB:S76436; NID:g243370; PIDN:AAB21121.1; PID:g243371
R;Dodge, G.R.; Kovalszky, I.; Chu, M.L.; Hassell, J.R.; McBride, O.W.; Yi, H.F.; Iozzo,
Genomics 10, 673-680, 1991
A;Title: Heparan sulfate proteoglycan of human colon: partial molecular cloning, cellula
A;Reference number: A40306; MUID:91365376; PMID:1679749
A;Accession: A40306
A;Molecule type: mRNA
A;Residues: 1018-1405,'G',1407-1409,'G',1411-1465 <DOD>
A;Cross-references: GB:M64283; NID:g184424; PIDN:AAA52699.1; PID:g184425
R;Heremans, A.; van der Schueren, B.; De Cock, B.; Paulsson, M.; Cassiman, J.J.; van den
J. Cell Biol. 109, 3199-3211, 1989
```

```
A;Title: Matrix-associated heparan sulfate proteoglycan: core protein-specific monoclona
anes.
A;Reference number: A33625; MUID:90078352; PMID:2687294
A;Accession: B33625
A;Molecule type: protein
A;Residues: 1379-1384,'X',1386-1388,'X',1390-1398 <HE2>
A;Accession: A33625
A;Molecule type: protein
A;Residues: 2166-2171,'X',2173-2175,'X',2177-2185 <HE3>
A;Note: peptide potentially matches four different regions of sequence shown
C;Genetics:
A;Gene: GDB:HSPG2
A;Cross-references: GDB:126372; OMIM:142461
A;Map position: 1p36.1-1p36.1
C;Superfamily: LDL receptor ligand-binding repeat homology; EGF homology; laminin G repe
C;Keywords: chondroitin sulfate proteoglycan; glycoprotein; heparan sulfate; transmembra
F;1-21/Domain: signal sequence #status predicted <SIG>
F;22-4391/Product: perlecan #status predicted <MAT>
F;22-193/Domain: I <DOM1>
F;194-530/Domain: II <DOM2>
F;199-234/Domain: LDL receptor ligand-binding repeat homology <LDL1>
F;285-319/Domain: LDL receptor ligand-binding repeat homology <LDL2>
F;325-359/Domain: LDL receptor ligand-binding repeat homology <LDL3>
F;368-403/Domain: LDL receptor ligand-binding repeat homology <LDL4>
F;531-1676/Domain: III <DOM3>
F;1159-1206/Domain: laminin-type EGF-like homology <LEG>
F;1563-1610/Domain: laminin-type EGF-like homology <EG7>
F;1613-1668/Domain: laminin-type EGF-like homology <LEG8>
F;1677-3686/Domain: IV <DOM4>
F;2007-2034/Domain: transmembrane #status predicted <TRM>
F;3687-4391/Domain: V <DOM5>
F;3845-3880/Domain: EGF homology <EGF1>
F;3888-3921/Domain: EGF homology <EGF>
F;3953-4106/Domain: laminin G repeat homology <LG2>
F;4147-4175/Domain: EGF homology <EGF2>
F;4149-4151/Region: motor neuron attachment (L-R-E) motif
F;4299-4301/Region: motor neuron attachment (L-R-E) motif
F;65,71,76/Binding site: heparan sulfate (Ser) (covalent) #status predicted
F;89,554,1755,2121,3072,3105,3279,3780,3836,4068/Binding site: carbohydrate (Asn) (coval
F;2995,3933,4179/Binding site: chondroitin sulfate (Ser) (covalent) #status predicted

  Query Match             100.0%;  Score 284;  DB 2;  Length 4391;
  Best Local Similarity   100.0%;  Pred. No. 6.4e-288;
  Matches  284;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 CERQPCQHGATCMPAGEYEFQCLCRDGFKGDLCEHEENPCQLREPCLHGGTCQGTRCLCL 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4108 CERQPCQHGATCMPAGEYEFQCLCRDGFKGDLCEHEENPCQLREPCLHGGTCQGTRCLCL 4167

Qy         61 PGFSGPRCQQGSGHGIAESDWHLEGSGGNDAPGQYGAYFHDDGFLAFPGHVFSRSLPEVP 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4168 PGFSGPRCQQGSGHGIAESDWHLEGSGGNDAPGQYGAYFHDDGFLAFPGHVFSRSLPEVP 4227

Qy        121 ETIELEVRTSTASGLLLWQGVEVGEAGQGKDFISLGLQDGHLVFRYQLGSGEARLVSEDP 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4228 ETIELEVRTSTASGLLLWQGVEVGEAGQGKDFISLGLQDGHLVFRYQLGSGEARLVSEDP 4287

Qy        181 INDGEWHRVTALREGRRGSIQVDGEELVSGRSPGPNVAVNAKGSVYIGGAPDVATLTGGR 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4288 INDGEWHRVTALREGRRGSIQVDGEELVSGRSPGPNVAVNAKGSVYIGGAPDVATLTGGR 4347

Qy        241 FSSGITGCVKNLVLHSARPGAPPPQPLDLQHRAQAGANTRPCPS 284
              ||||||||||||||||||||||||||||||||||||||||||||
Db       4348 FSSGITGCVKNLVLHSARPGAPPPQPLDLQHRAQAGANTRPCPS 4391


RESULT 2
S18252
heparan sulfate proteoglycan - mouse
N;Alternate names: perlecan
C;Species: Mus musculus (house mouse)
C;Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 05-Nov-1999
C;Accession: S18252; A31917; B31917; S66460
R;Noonan, D.M.; Fulle, A.; Valente, P.; Cai, S.; Horigan, E.; Sasaki, M.; Yamada, Y.; Ha
J. Biol. Chem. 266, 22939-22947, 1991
A;Title: The complete sequence of perlecan, a basement membrane heparan sulfate proteogl
adhesion molecule.
A;Reference number: S18252; MUID:92078153; PMID:1744087
A;Accession: S18252
A;Molecule type: mRNA
A;Residues: 1-3707 <NOO>
A;Cross-references: EMBL:M77174; NID:g200295; PIDN:AAA39911.1; PID:g200296
R;Noonan, D.M.; Horigan, E.A.; Ledbetter, S.R.; Vogeli, G.; Sasaki, M.; Yamada, Y.; Hass
J. Biol. Chem. 263, 16379-16387, 1988
A;Title: Identification of cDNA clones encoding different domains of the basement membra
A;Reference number: A92680; MUID:89034110; PMID:2972708
A;Accession: A31917
A;Molecule type: mRNA
A;Residues: 940-1601 <NO2>
A;Cross-references: GB:J04054; NID:g200252; PIDN:AAA39899.1; PID:g200253
A;Accession: B31917
A;Molecule type: mRNA
A;Residues: 1870-2600 <NO3>
A;Cross-references: GB:J04055; NID:g200300; PIDN:AAA39912.1; PID:g200301
R;Schulze, B.; Mann, K.; Battistutta, R.; Wiedemann, H.; Timpl, R.
Eur. J. Biochem. 231, 551-556, 1995
A;Title: Structural properties of recombinant domain III-3 of perlecan containing a glob
A;Reference number: S66460; MUID:95377282; PMID:7649154
A;Accession: S66460
A;Molecule type: protein
A;Residues: 1272-1274,'X',1276,'X',1278-1279 <SCH>
C;Superfamily: LDL receptor ligand-binding repeat homology; EGF homology; laminin G repe
C;Keywords: glycoprotein
F;199-234/Domain: LDL receptor ligand-binding repeat homology <LDL1>
F;285-319/Domain: LDL receptor ligand-binding repeat homology <LDL2>
F;325-359/Domain: LDL receptor ligand-binding repeat homology <LDL3>
F;368-403/Domain: LDL receptor ligand-binding repeat homology <LDL4>
F;764-811/Domain: laminin-type EGF-like homology <LEG>
F;1159-1206/Domain: laminin-type EGF-like homology <LEG7>
F;1563-1610/Domain: laminin-type EGF-like homology <EG7>
F;1613-1668/Domain: laminin-type EGF-like homology <LEG8>
F;3163-3198/Domain: EGF homology <EGF>
F;3270-3423/Domain: laminin G repeat homology <LG2>
F;3464-3492/Domain: EGF homology <EGF7>
F;1256,1891,2336,2394,2427/Binding site: carbohydrate (Asn) (covalent) #status predictec

  Query Match             9.9%;  Score 28;  DB 2;  Length 3707;
  Best Local Similarity   100.0%;  Pred. No. 5e-20;
  Matches   28;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        257 ARPGAPPPQPLDLQHRAQAGANTRPCPS 284
              ||||||||||||||||||||||||||||
Db       3680 ARPGAPPPQPLDLQHRAQAGANTRPCPS 3707


RESULT 3
A48825
Notch homolog Motch protein - mouse (fragment)
C;Species: Mus musculus (house mouse)
C;Date: 01-Dec-1993 #sequence_revision 18-Nov-1994 #text_change 08-Sep-2002
C;Accession: A48825
R;Reaume, A.G.; Conlon, R.A.; Zirngibl, R.; Yamaguchi, T.P.; Rossant, J.
Dev. Biol. 154, 377-387, 1992
A;Title: Expression analysis of a Notch homologue in the mouse embryo.
A;Reference number: A48825; MUID:93050801; PMID:1426644
A;Accession: A48825
A;Status: preliminary; not compared with conceptual translation
A;Molecule type: mRNA
A;Residues: 1-861 <REA>
A;Experimental source: embryo
A;Note: sequence extracted from NCBI backbone (NCBIP:119144)
C;Superfamily: notch protein; ankyrin repeat homology; EGF homology
F;26-57/Domain: EGF homology <EGF>
F;64-95/Domain: EGF homology <EGX1>
F;198-229/Domain: EGF homology <EGF2>
F;441-472/Domain: EGF homology <EGX2>
```

```
  Query Match             3.2%;  Score 9;  DB 2;  Length 861;
  Best Local Similarity   100.0%;  Pred. No. 1;
  Matches    9;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         45 PCLHGGTCQ 53
              |||||||||
Db        240 PCLHGGTCQ 248

RESULT 4
AGRT
agrin - rat
C;Species: Rattus norvegicus (Norway rat)
C;Date: 31-Mar-1993 #sequence_revision 31-Mar-1993 #text_change 17-Nov-2000
C;Accession: JH0399; A38856
R;Rupp, F.; Payan, D.G.; Magill-Solc, C.; Cowan, D.M.; Scheller, R.H.
Neuron 6, 811-823, 1991
A;Title: Structure and expression of a rat agrin.
A;Reference number: JH0399; MUID:91222570; PMID:1851019
A;Accession: JH0399
A;Molecule type: mRNA
A;Residues: 1-1779;1799-1959 <RUP>
A;Cross-references: GB:M64780; NID:g202798; PIDN:AAA40703.1; PID:g202800
A;Experimental source: embryonic spinal cord
A;Note: it is uncertain whether Met-1, Met-18, or Met-24 is the initiator
R;Rupp, F.; Oezcelik, T.; Linial, M.; Peterson, K.; Francke, U.; Scheller, R.
J. Neurosci. 12, 3535-3544, 1992
A;Title: Structure and chromosomal localization of the mammalian agrin gene.
A;Reference number: A38856; MUID:92407628; PMID:1326608
A;Accession: A38856
A;Molecule type: mRNA
A;Residues: 1780-1798 <RU2>
A;Cross-references: GB:S44194
C;Comment: This protein mediates the motor neuron-induced aggregation of acetylcholine r
C;Comment: 90% of rat embryonic transcripts encode the variant labeled below as form 3.
ylcholine receptor clustering activity.
C;Superfamily: agrin; EGF homology; Kazal proteinase inhibitor homology; laminin G repea
C;Keywords: alternative splicing; duplication; glycoprotein; neuromuscular junction
F;1-1959/Product: agrin, form 1 #status predicted <AG1>
F;1-1787,1799-1959/Product: agrin, form 4 #status predicted <AG4>
F;1-1779,1799-1959/Product: agrin, form 3 #status predicted <AG3>
F;1-1779,1788-1959/Product: agrin, form 5 #status predicted <AG5>
F;1-1143,1153-1959/Product: agrin, form 2 #status predicted <AG2>
F;22-50/Region: hydrophobic
F;88-137/Domain: Kazal proteinase inhibitor homology <KPI1>
F;163-212/Domain: Kazal proteinase inhibitor homology <KPI2>
F;236-284/Domain: Kazal proteinase inhibitor homology <KPI3>
F;307-356/Domain: Kazal proteinase inhibitor homology <KPI4>
F;381-429/Domain: Kazal proteinase inhibitor homology <KPI5>
F;446-494/Domain: Kazal proteinase inhibitor homology <KPI6>
F;511-559/Domain: Kazal proteinase inhibitor homology <KPI7>
F;540-542/Region: motor neuron attachment (L-R-E) motif
F;596-645/Domain: Kazal proteinase inhibitor homology <KPI8>
F;688-739/Domain: laminin-type EGF-like homology <LE1>
F;742-786/Domain: laminin-type EGF-like homology <LE2>
F;814-864/Domain: Kazal proteinase inhibitor homology <KPI9>
F;869-992/Region: serine/threonine-rich
F;1084-1086/Region: motor neuron attachment (L-R-E) motif
F;1147-1215/Region: serine/threonine-rich
F;1224-1257/Domain: EGF homology <EG1>
F;1287-1442/Domain: laminin G repeat homology <LG1>
F;1444-1476/Domain: EGF homology <EG2>
F;1483-1515/Domain: EGF homology <EG3>
F;1555-1706/Domain: laminin G repeat homology <LG2>
F;1713-1747/Domain: EGF homology <EG4>
F;1807-1959/Domain: laminin G repeat homology <LG3>
F;97-116,105-137,171-191,180-212,244-263,252-284,316-335,324-356,389-408,397-429,454-473
-1476,1483-1494,1488-1504,1506-1515/Disulfide bonds: #status predicted
F;145,672,827,957/Binding site: carbohydrate (Asn) (covalent) #status predicted

  Query Match             3.2%;  Score 9;  DB 1;  Length 1959;
  Best Local Similarity   100.0%;  Pred. No. 2.2;
```

```
  Matches    9;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         45 PCLHGGTCQ 53
              |||||||||
Db       1228 PCLHGGTCQ 1236

RESULT 5
S18188
notch protein homolog - rat
C;Species: Rattus norvegicus (Norway rat)
C;Date: 19-Feb-1994 #sequence_revision 10-Nov-1995 #text_change 02-Aug-2002
C;Accession: S18188
R;Weinmaster, G.; Roberts, V.J.; Lemke, G.
Development 113, 199-205, 1991
A;Title: A homolog of Drosophila Notch expressed during mammalian development.
A;Reference number: S18188; MUID:92111383; PMID:1764995
A;Accession: S18188
A;Molecule type: mRNA
A;Residues: 1-2531 <WEI>
A;Cross-references: EMBL:X57405; NID:g57634; PID:g57635
C;Superfamily: notch protein; ankyrin repeat homology; EGF homology
F;987-1018/Domain: EGF homology <EGF1>
F;1025-1056/Domain: EGF homology <EGF>
F;1233-1264/Domain: EGF homology <EGF2>
F;1917-1949/Domain: ankyrin repeat homology <AN1>
F;1950-1982/Domain: ankyrin repeat homology <AN2>
F;1984-2016/Domain: ankyrin repeat homology <AN3>
F;2017-2049/Domain: ankyrin repeat homology <AN4>
F;2050-2082/Domain: ankyrin repeat homology <AN5>

  Query Match             3.2%;  Score 9;  DB 2;  Length 2531;
  Best Local Similarity   100.0%;  Pred. No. 2.7;
  Matches    9;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         45 PCLHGGTCQ 53
              |||||||||
Db       1029 PCLHGGTCQ 1037

RESULT 6
A46019
notch-1 protein - mouse
N;Alternate names: motch protein
C;Species: Mus musculus (house mouse)
C;Date: 22-Sep-1993 #sequence_revision 18-Nov-1994 #text_change 07-Mar-2003
C;Accession: A46019; S25144; C49175; B46438; A46438; PH1569; S32109
R;del Amo, F.F.; Gendron-Maguire, M.; Swiatek, P.J.; Jenkins, N.A.; Copeland, N.G.; Grid
Genomics 15, 259-264, 1993
A;Title: Cloning, analysis, and chromosomal localization of Notch-1, a mouse homolog of
A;Reference number: A46019; MUID:93194170; PMID:8449489
A;Accession: A46019
A;Status: not compared with conceptual translation
A;Molecule type: nucleic acid
A;Residues: 1-2531 <DEL>
A;Cross-references: GB:Z11886; GB:S47228; NID:g288502; PIDN:CAA77941.1; PID:g288503
A;Note: sequence extracted from NCBI backbone (NCBIP:127318)
R;Franco del Amo, F.; Smith, D.E.; Swiatek, P.J.; Gendron-Maguire, M.; Greenspan, R.J.;
submitted to the EMBL Data Library, April 1992
A;Description: Expression pattern of Motch, a mouse homolog of Drosophila Notch, suggest
A;Reference number: S25144
A;Accession: S25144
A;Molecule type: mRNA
A;Residues: 1551-2108,'Q',2110-2114,'ALP',2118-2170 <FRA>
A;Cross-references: EMBL:Z11886
R;Lardelli, M.; Lendahl, U.
Exp. Cell Res. 204, 364-372, 1993
A;Title: Motch A and Motch B--two mouse Notch homologues coexpressed in a wide variety o
A;Reference number: A49175; MUID:93178563; PMID:8440332
A;Accession: C49175
A;Status: preliminary; nucleic acid sequence not shown
A;Molecule type: mRNA
A;Residues: 1161-1547 <LAR>
```

```
A;Cross-references: EMBL:X68278; NID:g287987; PIDN:CAA48339.1; PID:g287988
A;Experimental source: embryo
A;Note: sequence extracted from NCBI backbone (NCBIP:126159)
R;Kopan, R.; Weintraub, H.
J. Cell Biol. 121, 631-641, 1993
A;Title: Mouse notch: expression in hair follicles correlates with cell fate determinati
A;Reference number: A46438; MUID:93252998; PMID:8486742
A;Accession: B46438
A;Status: preliminary
A;Molecule type: nucleic acid
A;Residues: 1865-1932,'RR',1935-1937,'L',1938-1967,'I',1969-2044,'IE',2047-2052,'S',2054
A;Experimental source: embryo
A;Note: sequence extracted from NCBI backbone (NCBIN:131246, NCBIP:131247)
C;Comment: This protein has many EGF repeats and lin-12[#1172]/Notch repeats.
C;Comment: This protein is one of the neurogenic proteins controlling the decision betwe
C;Genetics:
A;Gene: notch-1
A;Map position: 2
A;Note: proximal region of chromosome 2
C;Superfamily: notch protein; ankyrin repeat homology; EGF homology
F;106-138/Domain: EGF homology <EGF1>
F;144-175/Domain: EGF homology <EG01>
F;222-254/Domain: EGF homology <EGF2>
F;261-292/Domain: EGF homology <EG02>
F;339-370/Domain: EGF homology <EG03>
F;416-449/Domain: EGF homology <EGF3>
F;456-487/Domain: EGF homology <EG04>
F;494-525/Domain: EGF homology <EG05>
F;532-563/Domain: EGF homology <EG06>
F;607-638/Domain: EGF homology <EG07>
F;682-713/Domain: EGF homology <EG08>
F;757-788/Domain: EGF homology <EG09>
F;795-826/Domain: EGF homology <EG10>
F;873-904/Domain: EGF homology <EG11>
F;911-942/Domain: EGF homology <EG12>
F;949-980/Domain: EGF homology <EG13>
F;987-1018/Domain: EGF homology <EG14>
F;1025-1056/Domain: EGF homology <EG15>
F;1063-1094/Domain: EGF homology <EG16>
F;1149-1180/Domain: EGF homology <EG17>
F;1187-1218/Domain: EGF homology <EG18>
F;1233-1264/Domain: EGF homology <EGF4>
F;1352-1383/Domain: EGF homology <EG19>
F;1391-1425/Domain: EGF homology <EGF>
F;1917-1948/Domain: ankyrin repeat homology <AN1>
F;1949-1981/Domain: ankyrin repeat homology <AN2>
F;1983-2015/Domain: ankyrin repeat homology <AN3>
F;2016-2048/Domain: ankyrin repeat homology <AN4>
F;2049-2081/Domain: ankyrin repeat homology <AN5>

  Query Match            3.2%;  Score 9;  DB 2;  Length 2531;
  Best Local Similarity  100.0%;  Pred. No. 2.7;
  Matches    9;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         45 PCLHGGTCQ 53
              |||||||||
Db       1029 PCLHGGTCQ 1037


RESULT 7
AH2015
sodium/solute symporter [imported] - Nostoc sp. (strain PCC 7120)
C;Species: Nostoc sp. PCC 7120
A;Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
C;Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 09-Dec-2002
C;Accession: AH2015
R;Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi
 Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S
DNA Res. 8, 205-213, 2001
A;Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Ana
A;Reference number: AB1807; MUID:21595285; PMID:11759840
A;Accession: AH2015
A;Status: preliminary
```

```
A;Molecule type: DNA
A;Residues: 1-497 <KUR>
A;Cross-references: GB:BA000019; PIDN:BAB78044.1; PID:g17135498; GSPDB:GN00179
A;Experimental source: strain PCC 7120
C;Genetics:
A;Gene: all1678

  Query Match            2.8%;  Score 8;  DB 2;  Length 497;
  Best Local Similarity  100.0%;  Pred. No. 7;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        130 STASGLLL 137
              ||||||||
Db        325 STASGLLL 332


RESULT 8
F70012
leucyl aminopeptidase homolog yuiE - Bacillus subtilis
C;Species: Bacillus subtilis
C;Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 03-Mar-2003
C;Accession: F70012
R;Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Berter
C.; Bron, S.; Brouillet, S.; Bruschi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Cho
A.; Ehrlich, S.D.; Emmerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari, E.
Nature 390, 249-256, 1997
A;Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Galler
iech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M.F.
Koetter, P.; Koningstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois,
A;Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Maueel
y, M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetelle
 Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadaie, Y.; Sato, T.; Scanlon,
A;Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Seror
akeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiyama,
T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K
A;Authors: Yoshikawa, H.F.; Zumstein, E.; Yoshikawa, H.; Danchin, A.
A;Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis.
A;Reference number: A69580; MUID:98044033; PMID:9384377
A;Accession: F70012
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-500 <KUN>
A;Cross-references: GB:Z99120; GB:AL009126; NID:g2635613; PIDN:CAB15195.1; PID:g2635702
A;Experimental source: strain 168
C;Genetics:
A;Gene: yuiE
C;Superfamily: Cytosol aminopeptidase

  Query Match            2.8%;  Score 8;  DB 2;  Length 500;
  Best Local Similarity  100.0%;  Pred. No. 7.1;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        232 DVATLTGG 239
              ||||||||
Db        367 DVATLTGG 374


RESULT 9
A49175
Motch B protein - mouse (fragment)
N;Alternate names: Notch homolog
C;Species: Mus musculus (house mouse)
C;Date: 21-Jan-1994 #sequence_revision 05-Jan-1996 #text_change 08-Sep-2002
C;Accession: A49175; PH1570; S32113
R;Lardelli, M.; Lendahl, U.
Exp. Cell Res. 204, 364-372, 1993
A;Title: Motch A and Motch B--two mouse Notch homologues coexpressed in a wide variety c
A;Reference number: A49175; MUID:93178563; PMID:8440332
A;Accession: A49175
A;Status: preliminary; nucleic acid sequence not shown
A;Molecule type: mRNA
A;Residues: 1-1203 <LAR>
A;Cross-references: EMBL:X68279; NID:g287989; PIDN:CAA48340.1; PID:g287990
```

```
A;Experimental source: embryo
A;Note: sequence extracted from NCBI backbone (NCBIP:126158)
C;Comment: This protein has many EGF repeats and lin-12/Notch repeats.
C;Comment: This protein is one of the neurogenic proteins controlling the decision betwe
C;Superfamily: notch protein; ankyrin repeat homology; EGF homology
F;143-174/Domain: EGF homology <EGX1>
F;482-513/Domain: EGF homology <EGF1>
F;560-591/Domain: EGF homology <EGF>
F;674-705/Domain: EGF homology <EGX2>
F;712-743/Domain: EGF homology <EGF3>
F;836-867/Domain: EGF homology <EGX3>

  Query Match             2.8%;  Score 8;  DB 2;  Length 1203;
  Best Local Similarity   100.0%;  Pred. No. 16;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          5 PCQHGATC 12
                ||||||||
Db        840 PCQHGATC 847

RESULT 10
S28764
neurocan precursor - rat
C;Species: Rattus norvegicus (Norway rat)
C;Date: 22-Nov-1993 #sequence_revision 01-Sep-1995 #text_change 04-Feb-2000
C;Accession: S28764
R;Rauch, U.; Karthikeyan, L.; Maurel, P.; Margolis, R.U.; Margolis, R.K.
J. Biol. Chem. 267, 19536-19547, 1992
A;Title: Cloning and primary structure of neurocan, a developmentally regulated, aggrega
A;Reference number: S28764; MUID:92406907; PMID:1326557
A;Accession: S28764
A;Molecule type: mRNA
A;Residues: 1-1257 <RAU>
A;Cross-references: EMBL:M97161; NID:g205649; PIDN:AAC37679.1; PID:g205650
C;Superfamily: aggrecan; C-type lectin homology; complement factor H repeat homology; EG
C;Keywords: chondroitin sulfate proteoglycan; glycoprotein
F;1-22/Domain: signal sequence #status predicted <SIG>
F;23-1257/Product: neurocan #status predicted <MAT>
F;176-253/Domain: link protein repeat homology <LNK1>
F;274-355/Domain: link protein repeat homology <LNK2>
F;364-366/Region: cell attachment (R-G-D) motif
F;953-984/Domain: EGF homology <EGF>
F;1029-1149/Domain: C-type lectin homology <LCH>
F;1156-1212/Domain: complement factor H repeat homology <FHD>
F;121,339,737,967,1164/Binding site: carbohydrate (Asn) (covalent) #status predicted
F;372,410/Binding site: chondroitin sulfate (Ser) (covalent) #status predicted
F;944/Binding site: chondroitin sulfate (Ser) (covalent) #status experimental

  Query Match             2.8%;  Score 8;  DB 2;  Length 1257;
  Best Local Similarity   100.0%;  Pred. No. 16;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         45 PCLHGGTC 52
                ||||||||
Db        957 PCLHGGTC 964

RESULT 11
S52781
neurocan - mouse
C;Species: Mus musculus (house mouse)
C;Date: 19-May-1995 #sequence_revision 21-Jul-1995 #text_change 04-Feb-2000
C;Accession: S52781
R;Rauch, U.; Forsberg, N.; Kulbe, G.; Arnold-Ammer, I.; Faessler, R.
submitted to the EMBL Data Library, February 1995
A;Description: Amino acid sequence of mouse neurocan and brevican and their different ex
A;Reference number: S52781
A;Accession: S52781
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-1268 <RAU>
A;Cross-references: EMBL:X84727; NID:g758629; PIDN:CAA59216.1; PID:g758630
```

```
C;Superfamily: aggrecan; C-type lectin homology; complement factor H repeat homology; EG
F;176-253/Domain: link protein repeat homology <LNK1>
F;274-355/Domain: link protein repeat homology <LNK2>
F;964-995/Domain: EGF homology <EGF>
F;1040-1160/Domain: C-type lectin homology <LCH>
F;1167-1223/Domain: complement factor H repeat homology <FHD>

  Query Match             2.8%;  Score 8;  DB 2;  Length 1268;
  Best Local Similarity   100.0%;  Pred. No. 16;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         45 PCLHGGTC 52
                ||||||||
Db        968 PCLHGGTC 975

RESULT 12
T43060
agrin - electric ray (Discopyge ommata) (fragment)
C;Species: Discopyge ommata
C;Date: 11-Jan-2000 #sequence_revision 11-Jan-2000 #text_change 17-Nov-2000
C;Accession: T43060
R;Smith, M.A.; Magill-Solc, C.; Rupp, F.; Yao, Y.M.M.; Schilling, J.W.; Snow, P.; McMaha
submitted to the EMBL Data Library, September 1992
A;Reference number: Z22308
A;Accession: T43060
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-1328 <SMI>
A;Cross-references: EMBL:L01423; NID:g213102; PID:g213103; PIDN:AAA49224.1
C;Superfamily: agrin; EGF homology; Kazal proteinase inhibitor homology; laminin G repea
C;Keywords: glycoprotein; neuromuscular junction

  Query Match             2.8%;  Score 8;  DB 2;  Length 1328;
  Best Local Similarity   100.0%;  Pred. No. 17;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         45 PCLHGGTC 52
                ||||||||
Db        616 PCLHGGTC 623

RESULT 13
MMHUMH
laminin alpha-2 chain - human (fragment)
N;Alternate names: laminin M chain; merosin heavy chain
C;Species: Homo sapiens (man)
C;Date: 30-Sep-1991 #sequence_revision 21-Aug-1998 #text_change 10-Dec-1999
C;Accession: PX0082; A35899; A38970; S14461
R;Hori, H.; Kanamori, T.; Mizuta, T.; Yamaguchi, N.; Liu, Y.; Nagai, Y.
J. Biochem. 116, 1212-1219, 1994
A;Title: Human laminin M chain: Epitope analysis of its monoclonal antibodies by immunos
A;Reference number: PX0082; MUID:95221315; PMID:7535762
A;Accession: PX0082
A;Molecule type: mRNA
A;Residues: 1-1751 <HOR>
A;Experimental source: placenta
R;Ehrig, K.; Leivo, I.; Argraves, W.S.; Ruoslahti, E.; Engvall, E.
Proc. Natl. Acad. Sci. U.S.A. 87, 3264-3268, 1990
A;Title: Merosin, a tissue-specific basement membrane protein, is a laminin-like protein
A;Reference number: A35899; MUID:90238994; PMID:2185464
A;Accession: A35899
A;Status: nucleic acid sequence not shown
A;Molecule type: mRNA
A;Residues: 'V',623-1751 <EHR1>
A;Cross-references: EMBL:M59832
A;Accession: A38970
A;Molecule type: protein
A;Residues: 1368-1384;1389-1406;1593-1607 <EHR2>
A;Note: the sequence from Fig. 1 is inconsistent with that from Fig. 2 in lacking 1599-V
R;Ehrig, K.; Leivo, I.; Argraves, S.W.; Ruoslahti, E.; Engvall, E.
submitted to the EMBL Data Library, December 1990
A;Description: The tissue-specific basement membrane protein merosin is a laminin-like p
```

```
A;Reference number: S14461
A;Accession: S14461
A;Molecule type: mRNA
A;Residues: 'V',623-1264,'R',1266-1751 <LEI>
A;Cross-references: EMBL:M59832; NID:g187520; PIDN:AAA63215.1; PID:g187521
C;Comment: This protein is a prominent component of the basement membrane that mediates
C;Genetics:
A;Gene: GDB:LAMA2; LAMM
A;Cross-references: GDB:132362; OMIM:156225
A;Map position: 6q22-6q23
C;Complex: Laminins are trimers of an alpha-type, a beta-type, and a gamma-type laminin
C;Function:
A;Description: interact with cells and with other basement membrane proteins to promote
C;Superfamily: laminin alpha-1 chain; laminin G repeat homology; laminin-type EGF-like h
C;Keywords: basement membrane; calcium binding; coiled coil; extracellular matrix; glycc
F;21-58/Domain: laminin-type EGF-like homology #status atypical <LE01>
F;61-107/Domain: laminin-type EGF-like homology <LE02>
F;110-165/Domain: laminin-type EGF-like homology <LE03>
F;168-212/Domain: laminin-type EGF-like homology <LE04>
F;527-567,1071-1300/Region: 3DM and 2D9 binding
F;811-972/Domain: laminin G repeat homology <LG1>
F;1005-1165/Domain: laminin G repeat homology <LG2>
F;1191-1354/Domain: laminin G repeat homology <LG3>
F;1430-1578/Domain: laminin G repeat homology <LG4>
F;1605-1751/Domain: laminin G repeat homology <LG5>
F;120,238,255,341,451,542,557,561,658,669,686,767,881,1001,1076,1119,1192,1199,1289,1509

  Query Match             2.8%;  Score 8;  DB 1;  Length 1751;
  Best Local Similarity   100.0%;  Pred. No. 22;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        122 TIELEVRT 129
              ||||||||
Db       1429 TIELEVRT 1436

RESULT 14
AGCH
agrin precursor - chicken
C;Species: Gallus gallus (chicken)
C;Date: 31-Mar-1993 #sequence_revision 31-Mar-1993 #text_change 17-Nov-2000
C;Accession: JH0591; A38857; B38857; I50692
R;Tsim, K.W.K.; Ruegg, M.A.; Escher, G.; Kroeger, S.; McMahan, U.J.
Neuron 8, 677-689, 1992
A;Title: cDNA that encodes active agrin.
A;Reference number: JH0591; MUID:92232297; PMID:1314620
A;Accession: JH0591
A;Molecule type: mRNA
A;Residues: 1-1955 <TSI>
A;Cross-references: GB:M94271; NID:g211120; PIDN:AAA48585.1; PID:g211121
A;Experimental source: brain
R;Ruegg, M.A.; Tsim, K.W.K.; Horton, S.E.; Kroeger, S.; Escher, G.; Gensch, E.M.; McMaha
Neuron 8, 691-699, 1992
A;Title: The agrin gene codes for a family of basal lamina proteins that differ in funct
A;Reference number: A38857; MUID:92232298; PMID:1314621
A;Contents: alternative splicing
A;Accession: A38857
A;Molecule type: mRNA
A;Residues: 1132-1783;1795-1955 <RU2>
A;Cross-references: GB:M97371
A;Accession: B38857
A;Molecule type: mRNA
A;Residues: 1221-1647;1652-1783;1794-1955 <RU3>
A;Cross-references: GB:M97372
A;Note: translation of the nucleotide sequence is not complete
R;Thomas, W.S.; O'Dowd, D.K.; Smith, M.A.
Dev. Biol. 158, 523-535, 1993
A;Title: Developmental expression and alternative splicing of chick agrin RNA.
A;Reference number: I50692; MUID:93345745; PMID:8393816
A;Accession: I50692
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 'SHLSNEIPA',1784-1795 <THO>
```

```
A;Cross-references: EMBL:U07271; NID:g459665; PIDN:AAA16788.1; PID:g459666
C;Comment: This protein mediates the motor neuron-induced aggregation of acetylcholine r
C;Comment: Alternative splicing produces two inactive proteins: agrin-related protein 1
C;Superfamily: agrin; EGF homology; Kazal proteinase inhibitor homology; laminin G repea
C;Keywords: alternative splicing; duplication; glycoprotein; neuromuscular junction
F;1-38/Domain: signal sequence #status predicted <SIG>
F;39-1955/Product: agrin #status predicted <MAT>
F;39-1783,1795-1955/Product: agrin-related protein 1 #status predicted <AG1>
F;39-1647,1652-1783,1794-1955/Product: agrin-related protein 2 #status predicted <AG2>
F;77-126/Domain: Kazal proteinase inhibitor homology <KPI1>
F;152-201/Domain: Kazal proteinase inhibitor homology <KPI2>
F;225-273/Domain: Kazal proteinase inhibitor homology <KPI3>
F;295-344/Domain: Kazal proteinase inhibitor homology <KPI4>
F;370-418/Domain: Kazal proteinase inhibitor homology <KPI5>
F;435-483/Domain: Kazal proteinase inhibitor homology <KPI6>
F;500-548/Domain: Kazal proteinase inhibitor homology <KPI7>
F;584-633/Domain: Kazal proteinase inhibitor homology <KPI8>
F;675-726/Domain: laminin-type EGF-like homology <LE1>
F;729-773/Domain: laminin-type EGF-like homology <LE2>
F;801-851/Domain: Kazal proteinase inhibitor homology <KPI9>
F;856-995/Region: serine/threonine-rich
F;1150-1219/Region: serine/threonine-rich
F;1233-1264/Domain: EGF homology <EG1>
F;1294-1448/Domain: laminin G repeat homology <LG1>
F;1429-1431/Region: motor neuron attachment (L-R-E) motif
F;1450-1482/Domain: EGF homology <EG2>
F;1489-1521/Domain: EGF homology <EG3>
F;1560-1711/Domain: laminin G repeat homology <LG2>
F;1718-1751/Domain: EGF homology <EG4>
F;1803-1955/Domain: laminin G repeat homology <LG3>
F;86-105,94-126,160-180,169-201,233-252,241-273,304-323,312-344,378-397,386-418,443-462,
1482,1489-1500,1494-1510,1512-1521/Disulfide bonds: #status predicted
F;390,659,764,814/Binding site: carbohydrate (Asn) (covalent) #status predicted

  Query Match             2.8%;  Score 8;  DB 1;  Length 1955;
  Best Local Similarity   100.0%;  Pred. No. 24;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         45 PCLHGGTC 52
              ||||||||
Db       1237 PCLHGGTC 1244

RESULT 15
S78549
notch3 protein - human
C;Species: Homo sapiens (man)
C;Date: 24-Jul-1998 #sequence_revision 24-Jul-1998 #text_change 08-Sep-2002
C;Accession: S78549; S71825
R;Joutel, A.; Tournier-Lasserve, E.
submitted to the EMBL Data Library, April 1997
A;Reference number: S78549
A;Accession: S78549
A;Molecule type: mRNA
A;Residues: 1-2321 <JOU1>
A;Cross-references: EMBL:U97669; NID:g2668591; PIDN:AAB91371.1; PID:g2668592
R;Joutel, A.; Corpechot, C.; Ducros, A.; Vahedi, K.; Chabriat, H.; Mouton, P.; Alamowitc
x, M.M.; Weissenbach, J.; Bach, J.F.; Bousser, M.G.; Tournier-Lasserve, E.
Nature 383, 707-710, 1996
A;Title: Notch3 mutations in CADASIL, a hereditary adult-onset condition causing stroke
A;Reference number: S71825; MUID:97032728; PMID:8878478
A;Accession: S71825
A;Status: nucleic acid sequence not shown
A;Molecule type: DNA
A;Residues: 67-113;138-194;268-333,'G',335-346;536-613;716-765;1240-1279;1815-1888 <JOU2
A;Cross-references: EMBL:U97669
C;Genetics:
A;Gene: notch3
A;Map position: 19p13.1
C;Function:
A;Description: may be involved in pathogenesis of CADASIL, causing a type of stroke and
C;Superfamily: notch protein; ankyrin repeat homology; EGF homology
C;Keywords: tandem repeat; transmembrane protein
```

```
F;123-155/Domain: EGF homology <EGX1>
F;162-194/Domain: EGF homology <EGF1>
F;240-271/Domain: EGF homology <EGX2>
F;318-349/Domain: EGF homology <EGF>
F;473-504/Domain: EGF homology <EGX3>
F;853-884/Domain: EGF homology <EGF3>
F;928-959/Domain: EGF homology <EGX4>
F;1838-1870/Domain: ankyrin repeat homology <AN1>
F;1871-1903/Domain: ankyrin repeat homology <AN2>
F;1905-1937/Domain: ankyrin repeat homology <AN3>
F;1938-1970/Domain: ankyrin repeat homology <AN4>
F;1971-2003/Domain: ankyrin repeat homology <AN5>

  Query Match             2.8%;  Score 8;  DB 2;  Length 2321;
  Best Local Similarity   100.0%;  Pred. No. 28;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         62 GFSGPRCQ 69
              ||||||||
Db       1237 GFSGPRCQ 1244


Search completed: March  9, 2004, 17:31:27
Job time : 10.7114 secs
```

```
                GenCore version 5.1.6
         Copyright (c) 1993 - 2004  Compugen Ltd.

OM protein - protein search, using sw model

Run on:         March  9, 2004, 17:21:48 ; Search time 6.86906 Seconds
                                           (without alignments)
                                           2152.832 Million cell updates/sec

Title:          US-10-006-011A-9
Perfect score:  284
Sequence:       1 CERQPCQHGATCMPAGEYEF..........QPLDLQHRAQAGANTRPCPS 284

Scoring table:  OLIGO
                Gapop 60.0 , Gapext 60.0

Searched:       141681 seqs, 52070155 residues

Word size :       0

Total number of hits satisfying chosen parameters:      141681

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :       SwissProt_42:*

      Pred. No. is the number of results predicted by chance to have a
      score greater than or equal to the score of the result being printed,
      and is derived by analysis of the total score distribution.

                                  SUMMARIES
                   %
Result            Query
  No.    Score   Match Length DB  ID                    Description
-----------------------------------------------------------------------------
   1      284   100.0   4391  1  PGBM_HUMAN            P98160 homo sapien
   2       28     9.9   3707  1  PGBM_MOUSE            Q05793 mus musculu
   3        9     3.2   1959  1  AGRI_RAT              P25304 rattus norv
   4        9     3.2   2531  1  NTC1_MOUSE            Q01705 mus musculu
   5        9     3.2   2531  1  NTC1_RAT              Q07008 rattus norv
   6        8     2.8    500  1  AMPA_BACSU            O32106 bacillus su
   7        8     2.8   1257  1  PGCN_RAT              P55067 rattus norv
   8        8     2.8   1268  1  PGCN_MOUSE            P55066 mus musculu
   9        8     2.8   1321  1  PGCN_HUMAN            O14594 homo sapien
  10        8     2.8   1328  1  AGRI_DISOM            Q90404 discopyge o
  11        8     2.8   1955  1  AGRI_CHICK            P31696 gallus gall
  12        8     2.8   2321  1  NTC3_HUMAN            Q9um47 homo sapien
  13        8     2.8   2470  1  NTC2_MOUSE            O35516 mus musculu
  14        8     2.8   2471  1  NTC2_HUMAN            Q04721 homo sapien
  15        8     2.8   2471  1  NTC2_RAT              Q9qw30 rattus norv
  16        8     2.8   3106  1  LMA2_MOUSE            Q60675 mus musculu
  17        8     2.8   3110  1  LMA2_HUMAN            P24043 homo sapien
  18        7     2.5    184  1  RS7_THEAC             Q9hly1 thermoplasm
  19        7     2.5    184  1  RS7_THEVO             Q97cd9 thermoplasm
  20        7     2.5    240  1  XYNC_STRLI            P26220 streptomyce
  21        7     2.5    244  1  YRT2_CAEEL            Q10045 caenorhabdi
  22        7     2.5    248  1  SCTM_HUMAN            Q8wvn6 homo sapien
  23        7     2.5    255  1  TFDC_ALCEU            P05403 alcaligenes
  24        7     2.5    273  1  RL2_TREPA             O83222 treponema p
  25        7     2.5    296  1  CYCG_RHOSH            Q53143 rhodobacter
  26        7     2.5    305  1  FMT_THETH             P43523 thermus the
  27        7     2.5    325  1  NAS2_ORYSA            Q9feg8 oryza sativ
  28        7     2.5    332  1  NAS1_ORYSA            Q9sxq7 oryza sativ
  29        7     2.5    348  1  NU2M_LATCH            O03166 latimeria c
  30        7     2.5    427  1  MFGM_BOVIN            Q95114 bos taurus
  31        7     2.5    429  1  DAD1_RALSO            Q8y0w7 ralstonia s
  32        7     2.5    445  1  AMPA_MYCPN            P75206 mycoplasma
  33        7     2.5    447  1  AMPA_MYCGE            P47631 mycoplasma
  34        7     2.5    468  1  AMPA_NEIMA            Q9jti8 neisseria m
  35        7     2.5    475  1  AMT2_ARATH            Q9m6n7 arabidopsis
  36        7     2.5    490  1  AMPA_XANAC            Q8pgr0 xanthomonas
  37        7     2.5    491  1  AMPA_HAEIN            P45334 haemophilus
  38        7     2.5    491  1  AMPA_XYLFA            Q9ph08 xylella fas
  39        7     2.5    491  1  AMPA_XYLFT            Q87f32 xylella fas
  40        7     2.5    493  1  AMPA_CLOPE            Q8xhi3 clostridium
  41        7     2.5    493  1  AMPA_XANCP            Q8pcr4 xanthomonas
  42        7     2.5    494  1  AMPA_PASMU            P57823 pasteurella
  43        7     2.5    499  1  AMPA_BUCAI            P57448 buchnera ap
  44        7     2.5    500  1  AMA1_SHEON            Q8ei85 shewanella
  45        7     2.5    500  1  AMPA_RICCN            Q92j85 rickettsia

                              ALIGNMENTS

RESULT 1
PGBM_HUMAN
ID   PGBM_HUMAN     STANDARD;      PRT;  4391 AA.
AC   P98160; Q16287; Q9H3V5;
DT   01-OCT-1996 (Rel. 34, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Basement membrane-specific heparan sulfate proteoglycan core
DE   protein precursor (HSPG) (Perlecan) (PLC).
GN   HSPG2.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=92112994; PubMed=1730768;
RA   Kallunki P., Tryggvason K.;
RT   "Human basement membrane heparan sulfate proteoglycan core protein: a
RT   467-kD protein containing multiple domains resembling elements of the
RT   low density lipoprotein receptor, laminin, neural cell adhesion
RT   molecules, and epidermal growth factor.";
RL   J. Cell Biol. 116:559-571(1992).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Colon, and Skin;
RX   MEDLINE=92235084; PubMed=1569102;
RA   Murdoch A.D., Dodge G.R., Cohen I., Tuan R.S., Iozzo R.V.;
RT   "Primary structure of the human heparan sulfate proteoglycan from
RT   basement membrane (HSPG2/perlecan). A chimeric molecule with multiple
RT   domains homologous to the low density lipoprotein receptor, laminin,
RT   neural cell adhesion molecules, and epidermal growth factor.";
RL   J. Biol. Chem. 267:8544-8557(1992).
RN   [3]
RP   SEQUENCE OF 22-4391 FROM N.A., AND VARIANT SJS1 TYR-1532.
RX   MEDLINE=20553141; PubMed=11101850;
RA   Nicole S., Davoine C.-S., Topaloglu H., Cattolico L., Barral D.,
RA   Beighton P., Ben-Hamida C., Hammouda H., Cruaud C., White P.S.,
RA   Samson D., Urtizberea J.A., Lehmann-Horn F., Weissenbach J.,
RA   Hentati F., Fontaine B.;
RT   "Perlecan, the major proteoglycan of basement membranes, is altered in
RT   patients with Schwartz-Jampel syndrome (chondrodystrophic myotonia).";
RL   Nat. Genet. 26:480-483(2000).
RN   [4]
RP   SEQUENCE OF 1016-1470 FROM N.A.
RC   TISSUE=Colon;
RX   MEDLINE=91365376; PubMed=1679749;
RA   Dodge G.R., Kovalszky I., Chu M.L., Hassell J.R., McBride O.W.,
RA   Yi H.F., Iozzo R.V.;
RT   "Heparan sulfate proteoglycan of human colon: partial molecular
RT   cloning, cellular expression, and mapping of the gene (HSPG2) to the
RT   short arm of human chromosome 1.";
RL   Genomics 10:673-680(1991).
RN   [5]
RP   SEQUENCE OF 890-1396 FROM N.A.
RC   TISSUE=Fibrosarcoma;
```

```
RX   MEDLINE=92120660; PubMed=1685141;
RA   Kallunki P., Eddy R.L., Byers M.G., Kestila M., Shows T.B.,
RA   Tryggvason K.;
RT   "Cloning of human heparan sulfate proteoglycan core protein,
RT   assignment of the gene (HSPG2) to 1p36.1-->p35 and identification of
RT   a BamHI restriction fragment length polymorphism.";
RL   Genomics 11:389-396(1991).
RN   [6]
RP   SEQUENCE OF 1-21 FROM N.A.
RX   MEDLINE=94052171; PubMed=8234307;
RA   Cohen I.R., Graessel S., Murdoch A.D., Iozzo R.V.;
RT   "Structural characterization of the complete human perlecan gene and
RT   its promoter.";
RL   Proc. Natl. Acad. Sci. U.S.A. 90:10404-10408(1993).
RN   [7]
RP   CARBOHYDRATE-LINKAGE SITE ASN-2121.
RX   MEDLINE=22660472; PubMed=12754519;
RA   Zhang H., Li X.-J., Martin D.B., Aebersold R.;
RT   "Identification and quantification of N-linked glycoproteins using
RT   hydrazide chemistry, stable isotope labeling and mass spectrometry.";
RL   Nat. Biotechnol. 21:660-666(2003).
CC   -!- FUNCTION: This protein is an integral component of basement
CC       membranes. It is responsible for the fixed negative electrostatic
CC       charge and is involved in the charge-selective ultrafiltration
CC       properties. It serves as an attachment substrate for cells.
CC   -!- SUBUNIT: Purified perlecan has a strong tendency to aggregate in
CC       dimers or stellate structures. It interacts with other basement
CC       membrane components such as laminin, prolargin and collagen type
CC       IV.
CC   -!- SUBCELLULAR LOCATION: Extracellular.
CC   -!- TISSUE SPECIFICITY: Found in the basement membranes.
CC   -!- PTM: CONTAINS THREE HEPARAN SULFATE CHAINS AS WELL AS N-LINKED
CC       AND O-LINKED OLIGOSACCHARIDES.
CC   -!- DISEASE: Defects in HSPG2 are the cause of Schwartz-Jampel
CC       syndrome (SJS1) [MIM:255800]; a rare autosomal recessive disorder
CC       characterized by permanent myotonia (prolonged failure of muscle
CC       relaxation) and skeletal dysplasia, resulting in reduced stature,
CC       kyphoscoliosis, bowing of the diaphyses and irregular epiphyses.
CC   -!- SIMILARITY: Contains 4 LDL-receptor class A domains.
CC   -!- SIMILARITY: Contains 11 laminin EGF-like domains.
CC   -!- SIMILARITY: Contains 3 laminin IV domains.
CC   -!- SIMILARITY: Contains 22 immunoglobulin-like C2-type domains.
CC   -!- SIMILARITY: Contains 3 laminin G-like domains.
CC   -!- SIMILARITY: Contains 4 EGF-like domains.
CC   -!- SIMILARITY: Contains 1 SEA domain.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; X62515; CAA44373.1; -.
DR   EMBL; M85289; AAA52700.1; -.
DR   EMBL; AL445795; CAC18534.1; -.
DR   EMBL; M64283; AAA52699.1; -.
DR   EMBL; S76436; AAB21121.2; -.
DR   EMBL; L22078; -; NOT_ANNOTATED_CDS.
DR   PIR; A38096; A38096.
DR   HSSP; P00740; 1EDM.
DR   Siena-2DPAGE; P98160; -.
DR   Genew; HGNC:5273; HSPG2.
DR   MIM; 142461; -.
DR   MIM; 255800; -.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR000742; EGF_2.
DR   InterPro; IPR006209; EGF_like.
DR   InterPro; IPR006210; IEGF.
DR   InterPro; IPR007110; Ig-like.
DR   InterPro; IPR003599; Ig.
DR   InterPro; IPR003598; Ig_c2.
DR   InterPro; IPR003596; Ig_v.
DR   InterPro; IPR000034; Laminin_B.
DR   InterPro; IPR002049; Laminin_EGF.
DR   InterPro; IPR001791; Laminin_G.
DR   InterPro; IPR002172; LDL_receptor_A.
DR   InterPro; IPR000082; SEA_domain.
DR   Pfam; PF00008; EGF; 4.
DR   Pfam; PF00047; ig; 22.
DR   Pfam; PF00052; laminin_B; 3.
DR   Pfam; PF00053; laminin_EGF; 7.
DR   Pfam; PF00054; laminin_G; 3.
DR   Pfam; PF00057; ldl_recept_a; 4.
DR   Pfam; PF01390; SEA; 1.
DR   PRINTS; PR00261; LDLRECEPTOR.
DR   ProDom; PD003031; Laminin_B; 3.
DR   SMART; SM00181; EGF; 15.
DR   SMART; SM00180; EGF_Lam; 12.
DR   SMART; SM00409; IG; 22.
DR   SMART; SM00408; IGc2; 21.
DR   SMART; SM00406; IGv; 7.
DR   SMART; SM00281; LamB; 3.
DR   SMART; SM00282; LamG; 3.
DR   SMART; SM00192; LDLa; 4.
DR   SMART; SM00200; SEA; 1.
DR   PROSITE; PS00022; EGF_1; 9.
DR   PROSITE; PS01186; EGF_2; 6.
DR   PROSITE; PS50026; EGF_3; 4.
DR   PROSITE; PS50835; IG_LIKE; 22.
DR   PROSITE; PS50025; LAM_G_DOMAIN; 3.
DR   PROSITE; PS01248; LAMININ_TYPE_EGF; 11.
DR   PROSITE; PS01209; LDLRA_1; 4.
DR   PROSITE; PS50068; LDLRA_2; 4.
DR   PROSITE; PS50024; SEA; 1.
KW   Signal; Basement membrane; Proteoglycan; Repeat; Glycoprotein;
KW   Heparan sulfate; Laminin EGF-like domain; Immunoglobulin domain;
KW   Extracellular matrix; EGF-like domain; Disease mutation.
FT   SIGNAL        1     21       POTENTIAL.
FT   CHAIN        22   4391       BASEMENT MEMBRANE-SPECIFIC HEPARAN
FT                                SULFATE PROTEOGLYCAN CORE PROTEIN.
FT   DOMAIN       80    194       SEA.
FT   DOMAIN      198    235       LDL-RECEPTOR CLASS A 1.
FT   DOMAIN      284    320       LDL-RECEPTOR CLASS A 2.
FT   DOMAIN      324    360       LDL-RECEPTOR CLASS A 3.
FT   DOMAIN      367    404       LDL-RECEPTOR CLASS A 4.
FT   DOMAIN      405    504       IG-LIKE C2-TYPE 1.
FT   DOMAIN      521    530       LAMININ EGF-LIKE 1 (N-TERMINAL).
FT   DOMAIN      531    730       LAMININ DOMAIN IV 1 (DOMAIN III A).
FT   DOMAIN      731    763       LAMININ EGF-LIKE 1 (C-TERMINAL).
FT   DOMAIN      764    813       LAMININ EGF-LIKE 2.
FT   DOMAIN      814    871       LAMININ EGF-LIKE 3.
FT   DOMAIN      879    923       LAMININ EGF-LIKE 4 (INCOMPLETE).
FT   DOMAIN      924    933       LAMININ EGF-LIKE 5 (N-TERMINAL).
FT   DOMAIN      934   1125       LAMININ DOMAIN IV 2 (DOMAIN III B).
FT   DOMAIN     1126   1158       LAMININ EGF-LIKE 5 (C-TERMINAL).
FT   DOMAIN     1159   1208       LAMININ EGF-LIKE 6.
FT   DOMAIN     1209   1265       LAMININ EGF-LIKE 7.
FT   DOMAIN     1275   1324       LAMININ EGF-LIKE 8.
FT   DOMAIN     1325   1334       LAMININ EGF-LIKE 9 (N-TERMINAL).
FT   DOMAIN     1335   1529       LAMININ DOMAIN IV 3 (DOMAIN III C).
FT   DOMAIN     1530   1562       LAMININ EGF-LIKE 9 (C-TERMINAL).
FT   DOMAIN     1563   1612       LAMININ EGF-LIKE 10.
FT   DOMAIN     1613   1670       LAMININ EGF-LIKE 11.
FT   DOMAIN     1677   1771       IG-LIKE C2-TYPE 2.
FT   DOMAIN     1772   1865       IG-LIKE C2-TYPE 3.
FT   DOMAIN     1866   1955       IG-LIKE C2-TYPE 4.
FT   DOMAIN     1956   2051       IG-LIKE C2-TYPE 5.
FT   DOMAIN     2052   2151       IG-LIKE C2-TYPE 6.
FT   DOMAIN     2152   2244       IG-LIKE C2-TYPE 7.
FT   DOMAIN     2245   2340       IG-LIKE C2-TYPE 8.
FT   DOMAIN     2341   2436       IG-LIKE C2-TYPE 9.
FT   DOMAIN     2437   2533       IG-LIKE C2-TYPE 10.
FT   DOMAIN     2534   2629       IG-LIKE C2-TYPE 11.
FT   DOMAIN     2630   2726       IG-LIKE C2-TYPE 12.
```

```
  Query Match             100.0%;  Score 284;  DB 1;  Length 4391;
  Best Local Similarity   100.0%;  Pred. No. 7.2e-283;
  Matches  284;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 CERQPCQHGATCMPAGEYEFQCLCRDGFKGDLCEHEENPCQLREPCLHGGTCQGTRCLCL 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4108 CERQPCQHGATCMPAGEYEFQCLCRDGFKGDLCEHEENPCQLREPCLHGGTCQGTRCLCL 4167

Qy         61 PGFSGPRCQQGSGHGIAESDWHLEGSGGNDAPGQYGAYFHDDGFLAFPGHVFSRSLPEVP 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4168 PGFSGPRCQQGSGHGIAESDWHLEGSGGNDAPGQYGAYFHDDGFLAFPGHVFSRSLPEVP 4227

Qy        121 ETIELEVRTSTASGLLLWQGVEVGEAGQGKDFISLGLQDGHLVFRYQLGSGEARLVSEDP 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4228 ETIELEVRTSTASGLLLWQGVEVGEAGQGKDFISLGLQDGHLVFRYQLGSGEARLVSEDP 4287

Qy        181 INDGEWHRVTALREGRRGSIQVDGEELVSGRSPGPNVAVNAKGSVYIGGAPDVATLTGGR 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4288 INDGEWHRVTALREGRRGSIQVDGEELVSGRSPGPNVAVNAKGSVYIGGAPDVATLTGGR 4347

Qy        241 FSSGITGCVKNLVLHSARPGAPPPQPLDLQHRAQAGANTRPCPS 284
              ||||||||||||||||||||||||||||||||||||||||||||
Db       4348 FSSGITGCVKNLVLHSARPGAPPPQPLDLQHRAQAGANTRPCPS 4391


RESULT 2
PGBM_MOUSE
ID   PGBM_MOUSE     STANDARD;      PRT;  3707 AA.
AC   Q05793;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Basement membrane-specific heparan sulfate proteoglycan core
DE   protein precursor (HSPG) (Perlecan) (PLC).
GN   HSPG2.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Melanoma;
RX   MEDLINE=92078153; PubMed=1744087;
RA   Noonan D.M., Fulle A., Valente P., Cai S., Horigan E., Sasaki M.,
RA   Yamada Y., Hassell J.R.;
RT   "The complete sequence of perlecan, a basement membrane heparan
RT   sulfate proteoglycan, reveals extensive similarity with laminin A
RT   chain, low density lipoprotein-receptor, and the neural cell adhesion
RT   molecule.";
RL   J. Biol. Chem. 266:22939-22947(1991).
RN   [2]
RP   SEQUENCE OF 940-1601 AND 1870-2600 FROM N.A., AND PARTIAL SEQUENCE.
RX   MEDLINE=89034110; PubMed=2972708;
RA   Noonan D.M., Horigan E.A., Ledbetter S.R., Vogeli G., Sasaki M.,
RA   Yamada Y., Hassell J.R.;
RT   "Identification of cDNA clones encoding different domains of the
RT   basement membrane heparan sulfate proteoglycan.";
RL   J. Biol. Chem. 263:16379-16387(1988).
CC   -!- FUNCTION: This protein is an integral component of basement
CC       membranes. It is responsible for the fixed negative electrostatic
CC       charge and is involved in the charge-selective ultrafiltration
CC       properties. It serves as an attachment substrate for cells.
CC   -!- SUBUNIT: Purified perlecan has a strong tendency to aggregate in
CC       dimers or stellate structures. It interacts with other basement
CC       membrane components such as laminin, prolargin and collagen type
CC       IV.
CC   -!- SUBCELLULAR LOCATION: Extracellular.
CC   -!- TISSUE SPECIFICITY: Found in the basement membranes.
CC   -!- PTM: CONTAINS THREE HEPARAN SULFATE CHAINS AS WELL AS N-LINKED
CC       AND O-LINKED OLIGOSACCHARIDES.
CC   -!- SIMILARITY: Contains 4 LDL-receptor class A domains.
CC   -!- SIMILARITY: Contains 11 laminin EGF-like domains.
CC   -!- SIMILARITY: Contains 3 laminin IV domains.
CC   -!- SIMILARITY: Contains 15 immunoglobulin-like C2-type domains.
CC   -!- SIMILARITY: Contains 3 laminin G-like domains.
CC   -!- SIMILARITY: Contains 1 EGF-like domain.
CC   -!- SIMILARITY: Contains 1 SEA domain.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; M77174; AAA39911.1; -.
DR   EMBL; J04054; AAA39899.1; -.
DR   EMBL; J04055; AAA39912.1; -.
DR   PIR; S18252; S18252.
DR   PDB; 1GL4; 28-NOV-01.
DR   MGD; MGI:96257; Hspg2.
DR   GO; GO:0005604; C:basement membrane; IDA.
DR   GO; GO:0008104; P:protein localization; IMP.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR000742; EGF_2.
DR   InterPro; IPR006209; EGF_like.
DR   InterPro; IPR007110; Ig-like.
DR   InterPro; IPR003598; Ig_c2.
DR   InterPro; IPR000034; Laminin_B.
DR   InterPro; IPR002049; Laminin_EGF.
DR   InterPro; IPR001791; Laminin_G.
DR   InterPro; IPR002172; LDL_receptor_A.
DR   InterPro; IPR000082; SEA_domain.
DR   Pfam; PF00008; EGF; 4.
DR   Pfam; PF00047; ig; 15.
DR   Pfam; PF00052; laminin_B; 3.
DR   Pfam; PF00053; laminin_EGF; 7.
DR   Pfam; PF00054; laminin_G; 3.
DR   Pfam; PF00057; ldl_recept_a; 4.
DR   Pfam; PF01390; SEA; 1.
DR   PRINTS; PR00261; LDLRECEPTOR.
DR   ProDom; PD003031; Laminin_B; 3.
DR   SMART; SM00180; EGF_Lam; 7.
DR   SMART; SM00408; IGc2; 14.
DR   SMART; SM00281; LamB; 3.
DR   SMART; SM00282; LamG; 3.
DR   SMART; SM00192; LDLa; 4.
DR   SMART; SM00200; SEA; 1.
DR   PROSITE; PS00022; EGF_1; 8.
DR   PROSITE; PS01186; EGF_2; 5.
DR   PROSITE; PS50026; EGF_3; 4.
DR   PROSITE; PS50835; IG_LIKE; 15.
DR   PROSITE; PS50025; LAM_G_DOMAIN; 3.
DR   PROSITE; PS01248; LAMININ_TYPE_EGF; 11.
DR   PROSITE; PS01209; LDLRA_1; 4.
DR   PROSITE; PS50068; LDLRA_2; 4.
DR   PROSITE; PS50024; SEA; 1.
KW   Signal; Basement membrane; Proteoglycan; Repeat; Glycoprotein;
KW   Heparan sulfate; Laminin EGF-like domain; Immunoglobulin domain;
KW   Extracellular matrix; EGF-like domain; 3D-structure.
FT   SIGNAL        1     21       POTENTIAL.
FT   CHAIN        22   3707       BASEMENT MEMBRANE-SPECIFIC HEPARAN
FT                                SULFATE PROTEOGLYCAN CORE PROTEIN.
FT   DOMAIN       80    194       SEA.
FT   DOMAIN      195    234       LDL-RECEPTOR CLASS A 1.
FT   DOMAIN      281    319       LDL-RECEPTOR CLASS A 2.
FT   DOMAIN      320    359       LDL-RECEPTOR CLASS A 3.
FT   DOMAIN      360    403       LDL-RECEPTOR CLASS A 4.
FT   DOMAIN      404    504       IG-LIKE C2-TYPE 1.
FT   DOMAIN      521    530       LAMININ EGF-LIKE 1 (N-TERMINAL).
FT   DOMAIN      531    730       LAMININ DOMAIN IV 1 (DOMAIN III A).
FT   DOMAIN      731    763       LAMININ EGF-LIKE 1 (C-TERMINAL).
FT   DOMAIN      764    813       LAMININ EGF-LIKE 2.
FT   DOMAIN      814    871       LAMININ EGF-LIKE 3.
```

```
FT   DOMAIN      879    923       LAMININ EGF-LIKE 4 (INCOMPLETE).
FT   DOMAIN      924    933       LAMININ EGF-LIKE 5 (N-TERMINAL).
FT   DOMAIN      934   1125       LAMININ DOMAIN IV 2 (DOMAIN III B).
FT   DOMAIN     1126   1158       LAMININ EGF-LIKE 5 (C-TERMINAL).
FT   DOMAIN     1159   1208       LAMININ EGF-LIKE 6.
FT   DOMAIN     1209   1265       LAMININ EGF-LIKE 7.
FT   DOMAIN     1275   1324       LAMININ EGF-LIKE 8.
FT   DOMAIN     1325   1334       LAMININ EGF-LIKE 9 (N-TERMINAL).
FT   DOMAIN     1335   1529       LAMININ DOMAIN IV 3 (DOMAIN III C).
FT   DOMAIN     1530   1562       LAMININ EGF-LIKE 9 (C-TERMINAL).
FT   DOMAIN     1563   1612       LAMININ EGF-LIKE 10.
FT   DOMAIN     1613   1670       LAMININ EGF-LIKE 11.
FT   DOMAIN     1677   1771       IG-LIKE C2-TYPE 2.
FT   DOMAIN     1772   1865       IG-LIKE C2-TYPE 3.
FT   DOMAIN     1866   1954       IG-LIKE C2-TYPE 4.
FT   DOMAIN     1955   2049       IG-LIKE C2-TYPE 5.
FT   DOMAIN     2050   2148       IG-LIKE C2-TYPE 6.
FT   DOMAIN     2149   2244       IG-LIKE C2-TYPE 7.
FT   DOMAIN     2245   2343       IG-LIKE C2-TYPE 8.
FT   DOMAIN     2344   2436       IG-LIKE C2-TYPE 9.
FT   DOMAIN     2437   2532       IG-LIKE C2-TYPE 10.
FT   DOMAIN     2533   2619       IG-LIKE C2-TYPE 11.
FT   DOMAIN     2620   2720       IG-LIKE C2-TYPE 12.
FT   DOMAIN     2721   2809       IG-LIKE C2-TYPE 13.
FT   DOMAIN     2810   2895       IG-LIKE C2-TYPE 14.
FT   DOMAIN     2896   2980       IG-LIKE C2-TYPE 15.
FT   DOMAIN     2984   3162       LAMININ G-LIKE 1.
FT   DOMAIN     3163   3241       EGF-LIKE.
FT   DOMAIN     3245   3425       LAMININ G-LIKE 2.
FT   DOMAIN     3518   3705       LAMININ G-LIKE 3.
FT   SITE         65     67       HEPARAN SULFATE (POTENTIAL).
FT   SITE         71     73       HEPARAN SULFATE (POTENTIAL).
FT   SITE         76     78       HEPARAN SULFATE (POTENTIAL).
FT   SITE       3615   3617       MEDIATES MOTOR NEURON ATTACHMENT
FT                                (POTENTIAL).
FT   DISULFID    199    212       BY SIMILARITY.
FT   DISULFID    206    225       BY SIMILARITY.
FT   DISULFID    219    234       BY SIMILARITY.
FT   DISULFID    285    297       BY SIMILARITY.
FT   DISULFID    292    310       BY SIMILARITY.
FT   DISULFID    304    319       BY SIMILARITY.
FT   DISULFID    325    337       BY SIMILARITY.
FT   DISULFID    332    350       BY SIMILARITY.
FT   DISULFID    344    359       BY SIMILARITY.
FT   DISULFID    368    381       BY SIMILARITY.
FT   DISULFID    375    394       BY SIMILARITY.
FT   DISULFID    388    403       BY SIMILARITY.
FT   DISULFID    428    479       BY SIMILARITY.
FT   DISULFID    764    773       BY SIMILARITY.
FT   DISULFID    766    780       BY SIMILARITY.
FT   DISULFID    783    792       BY SIMILARITY.
FT   DISULFID    795    811       BY SIMILARITY.
FT   DISULFID    814    829       BY SIMILARITY.
FT   DISULFID    816    839       BY SIMILARITY.
FT   DISULFID    842    851       BY SIMILARITY.
FT   DISULFID    854    869       BY SIMILARITY.
FT   DISULFID   1159   1168       BY SIMILARITY.
FT   DISULFID   1161   1175       BY SIMILARITY.
FT   DISULFID   1178   1187       BY SIMILARITY.
FT   DISULFID   1190   1206       BY SIMILARITY.
FT   DISULFID   1209   1224       BY SIMILARITY.
FT   DISULFID   1211   1234       BY SIMILARITY.
FT   DISULFID   1237   1246       BY SIMILARITY.
FT   DISULFID   1249   1263       BY SIMILARITY.
FT   DISULFID   1275   1287       BY SIMILARITY.
FT   DISULFID   1277   1293       BY SIMILARITY.
FT   DISULFID   1295   1304       BY SIMILARITY.
FT   DISULFID   1307   1322       BY SIMILARITY.
FT   DISULFID   1563   1572       BY SIMILARITY.
FT   DISULFID   1565   1579       BY SIMILARITY.
FT   DISULFID   1582   1591       BY SIMILARITY.
FT   DISULFID   1594   1610       BY SIMILARITY.
FT   DISULFID   1613   1628       BY SIMILARITY.
```

```
FT   DISULFID   1615   1638       BY SIMILARITY.
FT   DISULFID   1641   1650       BY SIMILARITY.
FT   DISULFID   1653   1668       BY SIMILARITY.
FT   DISULFID   1792   1839       BY SIMILARITY.
FT   DISULFID   1886   1932       BY SIMILARITY.
FT   DISULFID   1976   2021       BY SIMILARITY.
FT   DISULFID   2073   2118       BY SIMILARITY.
FT   DISULFID   2170   2215       BY SIMILARITY.
FT   DISULFID   2268   2313       BY SIMILARITY.

  Query Match             9.9%;  Score 28;  DB 1;  Length 3707;
  Best Local Similarity   100.0%;  Pred. No. 8e-20;
  Matches   28;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        257 ARPGAPPPQPLDLQHRAQAGANTRPCPS 284
              ||||||||||||||||||||||||||||
Db       3680 ARPGAPPPQPLDLQHRAQAGANTRPCPS 3707


RESULT 3
AGRI_RAT
ID   AGRI_RAT       STANDARD;      PRT;  1959 AA.
AC   P25304; Q63034;
DT   01-MAY-1992 (Rel. 22, Created)
DT   01-JUL-1993 (Rel. 26, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Agrin precursor.
GN   AGRN.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE OF 1-1779 AND 1799-1959 FROM N.A.
RC   TISSUE=Embryonic spinal cord;
RX   MEDLINE=91222570; PubMed=1851019;
RA   Rupp F., Payan D.G., Magill-Solc C., Cowan D.M., Scheller R.H.;
RT   "Structure and expression of a rat agrin.";
RL   Neuron 6:811-823(1991).
RN   [2]
RP   SEQUENCE OF 1777-1801 FROM N.A.
RX   MEDLINE=92407628; PubMed=1326608;
RA   Rupp F., Oezcelik T., Linial M., Peterson K., Francke U., Scheller R.;
RT   "Structure and chromosomal localization of the mammalian agrin gene.";
RL   J. Neurosci. 12:3535-3544(1992).
CC   -!- FUNCTION: Component of the basal lamina that causes the
CC       aggregation of acetylcholine receptors and acetylcholine-esterase
CC       on the surface of muscle fibers of the neuromuscular junction.
CC   -!- SUBUNIT: Binds to laminin.
CC   -!- SUBCELLULAR LOCATION: Synaptic basal lamina at the neuromuscular
CC       junction.
CC   -!- ALTERNATIVE PRODUCTS:
CC       Event=Alternative splicing; Named isoforms=5;
CC         Comment=Additional isoforms seem to exist. Isoforms differ in
CC         their acetylcholine receptor clustering activity;
CC       Name=1;
CC         IsoId=P25304-1; Sequence=Displayed;
CC       Name=2;
CC         IsoId=P25304-2; Sequence=VSP_001365;
CC       Name=3;
CC         IsoId=P25304-3; Sequence=VSP_001366;
CC       Name=4;
CC         IsoId=P25304-4; Sequence=VSP_001367;
CC       Name=5;
CC         IsoId=P25304-5; Sequence=VSP_001368;
CC   -!- TISSUE SPECIFICITY: Embryonic nervous system and muscle.
CC   -!- DEVELOPMENTAL STAGE: More abundant early in development.
CC   -!- PTM: Contains heparan sulfate chains as well as N-linked and O-
CC       linked oligosaccharides (By similarity).
CC   -!- SIMILARITY: Contains 9 Kazal-like domains.
CC   -!- SIMILARITY: Contains 2 laminin EGF-like domains.
CC   -!- SIMILARITY: Contains 4 EGF-like domains.
CC   -!- SIMILARITY: Contains 1 SEA domain.
```

```
CC   -!- SIMILARITY: Contains 3 laminin G-like domains.
CC   -!- CAUTION: It is uncertain whether Met-1, Met-18 or Met-24 is the
CC       initiator.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; M64780; AAA40703.1; -.
DR   EMBL; M64780; AAA40702.1; ALT_INIT.
DR   EMBL; S44194; AAB23326.1; -.
DR   PIR; JH0399; AGRT.
DR   HSSP; P00740; 1EDM.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR006209; EGF_like.
DR   InterPro; IPR003645; FolN.
DR   InterPro; IPR002350; kazal.
DR   InterPro; IPR002049; Laminin_EGF.
DR   InterPro; IPR001791; Laminin_G.
DR   InterPro; IPR000082; SEA_domain.
DR   Pfam; PF00008; EGF; 4.
DR   Pfam; PF00050; kazal; 9.
DR   Pfam; PF00053; laminin_EGF; 2.
DR   Pfam; PF00054; laminin_G; 3.
DR   Pfam; PF01390; SEA; 1.
DR   PRINTS; PR00011; EGFLAMININ.
DR   SMART; SM00180; EGF_Lam; 2.
DR   SMART; SM00274; FOLN; 8.
DR   SMART; SM00280; KAZAL; 9.
DR   SMART; SM00282; LamG; 3.
DR   SMART; SM00200; SEA; 1.
DR   PROSITE; PS00022; EGF_1; 6.
DR   PROSITE; PS01186; EGF_2; 1.
DR   PROSITE; PS50026; EGF_3; 4.
DR   PROSITE; PS50025; LAM_G_DOMAIN; 3.
DR   PROSITE; PS01248; LAMININ_TYPE_EGF; 1.
DR   PROSITE; PS50024; SEA; 1.
KW   Glycoprotein; EGF-like domain; Repeat; Alternative splicing; Signal;
KW   Laminin EGF-like domain; Proteoglycan; Heparan sulfate.
FT   SIGNAL        1     29       POTENTIAL.
FT   CHAIN        30   1959       AGRIN.
FT   DOMAIN       65    137       KAZAL-LIKE 1.
FT   DOMAIN      141    212       KAZAL-LIKE 2.
FT   DOMAIN      213    284       KAZAL-LIKE 3.
FT   DOMAIN      287    356       KAZAL-LIKE 4.
FT   DOMAIN      361    429       KAZAL-LIKE 5.
FT   DOMAIN      430    494       KAZAL-LIKE 6.
FT   DOMAIN      495    559       KAZAL-LIKE 7.
FT   DOMAIN      563    645       KAZAL-LIKE 8.
FT   DOMAIN      688    741       LAMININ EGF-LIKE 1.
FT   DOMAIN      742    788       LAMININ EGF-LIKE 2.
FT   DOMAIN      794    864       KAZAL-LIKE 9.
FT   DOMAIN     1023   1145       SEA.
FT   DOMAIN     1220   1258       EGF-LIKE 1.
FT   DOMAIN     1263   1439       LAMININ G-LIKE 1.
FT   DOMAIN     1440   1477       EGF-LIKE 2.
FT   DOMAIN     1479   1516       EGF-LIKE 3.
FT   DOMAIN     1526   1708       LAMININ G-LIKE 2.
FT   DOMAIN     1709   1748       EGF-LIKE 4.
FT   DOMAIN     1784   1956       LAMININ G-LIKE 3.
FT   DOMAIN      869    992       SER/THR-RICH.
FT   DOMAIN     1147   1215       SER/THR-RICH.
FT   DISULFID     97    116       POTENTIAL.
FT   DISULFID    105    137       POTENTIAL.
FT   DISULFID    171    191       POTENTIAL.
FT   DISULFID    180    212       POTENTIAL.
FT   DISULFID    244    263       POTENTIAL.
FT   DISULFID    252    284       POTENTIAL.
FT   DISULFID    316    335       POTENTIAL.
```

```
FT   DISULFID    324    356       POTENTIAL.
FT   DISULFID    389    408       POTENTIAL.
FT   DISULFID    397    429       POTENTIAL.
FT   DISULFID    454    473       POTENTIAL.
FT   DISULFID    462    494       POTENTIAL.
FT   DISULFID    518    538       POTENTIAL.
FT   DISULFID    527    559       POTENTIAL.
FT   DISULFID    604    624       POTENTIAL.
FT   DISULFID    613    645       POTENTIAL.
FT   DISULFID    688    700       BY SIMILARITY.
FT   DISULFID    690    707       BY SIMILARITY.
FT   DISULFID    709    718       BY SIMILARITY.
FT   DISULFID    721    739       BY SIMILARITY.
FT   DISULFID    742    754       BY SIMILARITY.
FT   DISULFID    744    761       BY SIMILARITY.
FT   DISULFID    763    772       BY SIMILARITY.
FT   DISULFID    775    786       BY SIMILARITY.
FT   DISULFID    823    843       POTENTIAL.
FT   DISULFID    832    864       POTENTIAL.
FT   DISULFID   1224   1235       BY SIMILARITY.
FT   DISULFID   1229   1246       BY SIMILARITY.
FT   DISULFID   1248   1257       BY SIMILARITY.
FT   DISULFID   1444   1455       POTENTIAL.
FT   DISULFID   1449   1465       POTENTIAL.
FT   DISULFID   1467   1476       POTENTIAL.
FT   DISULFID   1483   1494       BY SIMILARITY.
FT   DISULFID   1488   1504       BY SIMILARITY.
FT   DISULFID   1506   1515       BY SIMILARITY.
FT   DISULFID   1713   1727       BY SIMILARITY.
FT   DISULFID   1721   1736       BY SIMILARITY.
FT   DISULFID   1738   1747       BY SIMILARITY.
FT   CARBOHYD    145    145       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    672    672       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    827    827       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    957    957       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   VARSPLIC   1144   1152       Missing (in isoform 2).
FT                                /FTId=VSP_001365.
FT   VARSPLIC   1780   1798       Missing (in isoform 3).
FT                                /FTId=VSP_001366.
FT   VARSPLIC   1788   1798       Missing (in isoform 4).
FT                                /FTId=VSP_001367.
FT   VARSPLIC   1780   1787       Missing (in isoform 5).
FT                                /FTId=VSP_001368.
FT   VARIANT     314    314       V -> VTCD (IN A VARIANT).
SQ   SEQUENCE   1959 AA;  208645 MW;  7FEFDFDAFF89CC31 CRC64;

  Query Match             3.2%;  Score 9;  DB 1;  Length 1959;
  Best Local Similarity   100.0%;  Pred. No. 1.5;
  Matches    9;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         45 PCLHGGTCQ 53
              |||||||||
Db       1228 PCLHGGTCQ 1236


RESULT 4
NTC1_MOUSE
ID   NTC1_MOUSE     STANDARD;      PRT;  2531 AA.
AC   Q01705; Q06007; Q61905; Q99JC2; Q9QW58; Q9R0X7;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-FEB-1996 (Rel. 33, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Neurogenic locus notch homolog protein 1 precursor (Notch 1) (Motch A)
DE   (mT14) (p300).
GN   NOTCH1 OR MOTCH.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A. (ISOFORM 1).
RC   TISSUE=Embryo;
RX   MEDLINE=93194170; PubMed=8449489;
```

```
RA   Franco del Amo F., Gendron-Maguire M., Swiatek P.J., Jenkins N.A.,
RA   Copeland N.G., Gridley T.;
RT   "Cloning, analysis, and chromosomal localization of Notch-1, a mouse
RT   homolog of Drosophila Notch.";
RL   Genomics 15:259-264(1993).
RN   [2]
RP   SEQUENCE OF 731-1899 FROM N.A. (ISOFORM 2), AND DEVELOPMENTAL STAGE.
RC   STRAIN=CD-1; TISSUE=Embryo;
RX   MEDLINE=93050801; PubMed=1426644;
RA   Reaume A.G., Conlon R.A., Zirngibl R., Yamaguchi T.P., Rossant J.;
RT   "Expression analysis of a Notch homologue in the mouse embryo.";
RL   Dev. Biol. 154:377-387(1992).
RN   [3]
RP   SEQUENCE OF 1551-1647 FROM N.A. (ISOFORM 1), AND DEVELOPMENTAL STAGE.
RC   TISSUE=Embryo;
RX   MEDLINE=93048835; PubMed=1425352;
RA   Franco del Amo F., Smith D.E., Swiatek P.J., Gendron-Maguire M.,
RA   Greenspan R.J., McMahon A.P., Gridley T.;
RT   "Expression pattern of Motch, a mouse homolog of Drosophila Notch,
RT   suggests an important role in early postimplantation mouse
RT   development.";
RL   Development 115:737-744(1992).
RN   [4]
RP   SEQUENCE OF 1161-1547 FROM N.A.
RC   STRAIN=C57BL/6 X CBA; TISSUE=Embryo;
RX   MEDLINE=93178563; PubMed=8440332;
RA   Lardelli M., Lendahl U.;
RT   "Motch A and Motch B-two mouse Notch homologues coexpressed in a
RT   wide variety of tissues.";
RL   Exp. Cell Res. 204:364-372(1993).
RN   [5]
RP   SEQUENCE OF 1659-1673 FROM N.A.
RX   MEDLINE=99364499; PubMed=10437788;
RA   Lee J.S., Ishimoto A., Yanagawa S.I.;
RT   "Murine leukemia provirus-mediated activation of the Notch1 gene leads
RT   to induction of HES-1 in a mouse T lymphoma cell line, DL-3.";
RL   FEBS Lett. 455:276-280(1999).
RN   [6]
RP   SEQUENCE OF 1950-2201 FROM N.A.
RX   MEDLINE=98029496; PubMed=9384671;
RA   Messerle M., Follo M., Nehls M., Eggert H., Boehm T.;
RT   "Dynamic changes in gene expression during in vitro differentiation of
RT   mouse embryonic stem cells.";
RL   Cytokines Cell. Mol. Ther. 1:139-143(1995).
RN   [7]
RP   SEQUENCE OF 1655-1659, CLEAVAGE BY FURIN-LIKE CONVERTASE, AND
RP   MUTAGENESIS OF 1651-ARG--ARG-1654.
RX   MEDLINE=98318619; PubMed=9653148;
RA   Logeat F., Bessia C., Brou C., LeBail O., Jarriault S., Seidah N.G.,
RA   Israel A.;
RT   "The Notch1 receptor is cleaved constitutively by a furin-like
RT   convertase.";
RL   Proc. Natl. Acad. Sci. U.S.A. 95:8108-8112(1998).
RN   [8]
RP   PARTIAL SEQUENCE, AND POST-TRANSLATIONAL PROCESSING.
RX   MEDLINE=21523956; PubMed=11518718;
RA   Saxena M.T., Schroeter E.H., Mumm J.S., Kopan R.;
RT   "Murine notch homologs (N1-4) undergo presenilin-dependent
RT   proteolysis.";
RL   J. Biol. Chem. 276:40268-40273(2001).
RN   [9]
RP   POST-TRANSLATIONAL PROCESSING.
RX   MEDLINE=21374376; PubMed=11459941;
RA   Mizutani T., Taniguchi Y., Aoki T., Hashimoto N., Honjo T.;
RT   "Conservation of the biochemical mechanisms of signal transduction
RT   among mammalian Notch family members.";
RL   Proc. Natl. Acad. Sci. U.S.A. 98:9026-9031(2001).
RN   [10]
RP   INTERACTION WITH DTX1 AND DTX2.
RX   MEDLINE=21123790; PubMed=11226752;
RA   Kishi N., Tang Z., Maeda Y., Hirai A., Mo R., Ito M., Suzuki S.,
RA   Nakao K., Kinoshita T., Kadesch T., Hui C.-C., Artavanis-Tsakonas S.,
RA   Okano H., Matsuno K.;
```

```
RT   "Murine homologs of deltex define a novel gene family involved in
RT   vertebrate Notch signaling and neurogenesis.";
RL   Int. J. Dev. Neurosci. 19:21-35(2001).
CC   -!- FUNCTION: Functions as a receptor for membrane-bound ligands
CC       Jagged1, Jagged2 and Delta1 to regulate cell-fate determination.
CC       Upon ligand activation through the released notch intracellular
CC       domain (NICD) it forms a transcriptional activator complex with
CC       RBP-J kappa and activates genes of the enhancer of split locus.
CC       Affects the implementation of differentiation, proliferation and
CC       apoptotic programs (By similarity). May play an essential role in
CC       postimplantation development, probably in some aspect of cell
CC       specification and/or differentiation. May be involved in mesoderm
CC       development, somite formation and neurogenesis. Involved in the
CC       maturation of both CD4+ and CD8+ cells in the thymus.
CC   -!- SUBUNIT: Heterodimer of a C-terminal fragment N(TM) and a N-
CC       terminal fragment N(EC) which are probably linked by disulfide
CC       bonds. Interacts with DTX1 and DTX2.
CC   -!- SUBCELLULAR LOCATION: Type I membrane protein. Following
CC       proteolytical processing NICD is translocated to the nucleus.
CC   -!- ALTERNATIVE PRODUCTS:
CC       Event=Alternative splicing; Named isoforms=2;
CC       Name=1;
CC         IsoId=Q01705-1; Sequence=Displayed;
CC       Name=2;
CC         IsoId=Q01705-2; Sequence=VSP_001402, VSP_001403, VSP_001404;
CC         Note=No experimental confirmation available;
CC   -!- TISSUE SPECIFICITY: Highly expressed in the brain, lung and
CC       thymus. Expressed at lower levels in the spleen, bone-marrow,
CC       spinal cord, eyes, mammary gland, liver, intestine, skeletal
CC       muscle, kidney and heart.
CC   -!- DEVELOPMENTAL STAGE: First detected in the mesoderm at 7.5 dpc By
CC       8.5 dpc highly expressed in presomitic mesoderm, mesenchyme and
CC       endothelial cells, while much lower levels are seen in the
CC       neuroepithelium. Between 9.5-10.5 dpc expressed at high levels in
CC       the neuroepithelium. At 13.5 dpc expressed in the surface
CC       ectoderm, eye and developing whisker follicles.
CC   -!- PTM: Synthesized in the endoplasmic reticulum as an inactive form
CC       which is proteolytically cleaved by a furin-like convertase in the
CC       trans-Golgi network before it reaches the plasma membrane to yield
CC       an active, ligand-accessible form. Cleavage results in a C-
CC       terminal fragment N(TM) and a N-terminal fragment N(EC). Following
CC       ligand binding, it is cleaved by TNF-alpha converting enzyme
CC       (TACE) to yield a membrane-associated intermediate fragment called
CC       notch extracellular truncation (NEXT). This fragment is then
CC       cleaved by presenilin dependent gamma-secretase to release a
CC       notch-derived peptide containing the intracellular domain (NICD)
CC       from the membrane.
CC   -!- PTM: Phosphorylated.
CC   -!- SIMILARITY: Belongs to the NOTCH family.
CC   -!- SIMILARITY: Contains 36 EGF-like domains.
CC   -!- SIMILARITY: Contains 3 Lin/Notch repeats.
CC   -!- SIMILARITY: Contains 5 ANK repeats.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; Z11886; CAA77941.1; -.
DR   EMBL; L02613; AAK14898.1; -.
DR   EMBL; X68278; CAA48339.1; -.
DR   EMBL; AJ238029; CAB40733.1; -.
DR   EMBL; X82562; CAA57909.1; -.
DR   PIR; A46019; A46019.
DR   PIR; B49175; B49175.
DR   HSSP; P00740; 1EDM.
DR   MGD; MGI:97363; Notch1.
DR   GO; GO:0005887; C:integral to plasma membrane; IC.
DR   GO; GO:0005515; F:protein binding; IPI.
DR   GO; GO:0030154; P:cell differentiation; IMP.
```

```
DR   GO; GO:0007386; P:compartment specification; IMP.
DR   GO; GO:0007219; P:N signaling pathway; IC.
DR   GO; GO:0045944; P:positive regulation of transcription from P. . .; IDA.
DR   InterPro; IPR002110; ANK.
DR   InterPro; IPR000152; Asx_hydroxyl_S.
DR   InterPro; IPR000742; EGF_2.
DR   InterPro; IPR001881; EGF_Ca.
DR   InterPro; IPR001438; EGF_II.
DR   InterPro; IPR006209; EGF_like.
DR   InterPro; IPR002049; Laminin_EGF.
DR   InterPro; IPR008297; Notch.
DR   InterPro; IPR000800; Notch_dom.
DR   Pfam; PF00023; ank; 7.
DR   Pfam; PF00008; EGF; 35.
DR   Pfam; PF00066; notch; 3.
DR   PIRSF; PIRSF002279; Notch; 1.
DR   PRINTS; PR00010; EGFBLOOD.
DR   PRINTS; PR00011; EGFLAMININ.
DR   PRINTS; PR01452; NOTCH.
DR   SMART; SM00248; ANK; 6.
DR   SMART; SM00179; EGF_CA; 24.
DR   SMART; SM00004; NL; 2.
DR   PROSITE; PS50297; ANK_REP_REGION; 1.
DR   PROSITE; PS50088; ANK_REPEAT; 2.
DR   PROSITE; PS00010; ASX_HYDROXYL; 22.
DR   PROSITE; PS00022; EGF_1; 34.
DR   PROSITE; PS01186; EGF_2; 27.
DR   PROSITE; PS50026; EGF_3; 36.
DR   PROSITE; PS01187; EGF_CA; 21.
KW   Receptor; Transcription regulation; Activator; Differentiation;
KW   Developmental protein; Repeat; ANK repeat; EGF-like domain;
KW   Transmembrane; Glycoprotein; Signal; Phosphorylation;
KW   Alternative splicing.
FT   SIGNAL        1     18       POTENTIAL.
FT   CHAIN        19   2531       NEUROGENIC LOCUS NOTCH HOMOLOG PROTEIN 1.
FT   CHAIN      1711   2531       NOTCH EXTRACELLULAR TRUNCATION.
FT   CHAIN      1744   2531       NOTCH INTRACELLULAR DOMAIN.
FT   DOMAIN       19   1725       EXTRACELLULAR (POTENTIAL).

  Query Match             3.2%;  Score 9;  DB 1;  Length 2531;
  Best Local Similarity   100.0%;  Pred. No. 1.9;
  Matches    9;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         45 PCLHGGTCQ 53
              |||||||||
Db       1029 PCLHGGTCQ 1037


RESULT 5
NTC1_RAT
ID   NTC1_RAT       STANDARD;      PRT;  2531 AA.
AC   Q07008;
DT   01-NOV-1995 (Rel. 32, Created)
DT   15-JUL-1999 (Rel. 38, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Neurogenic locus notch homolog protein 1 precursor (Notch 1).
GN   NOTCH1.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Schwann cell;
RX   MEDLINE=92111383; PubMed=1764995;
RA   Weinmaster G., Roberts V.J., Lemke G.;
RT   "A homolog of Drosophila Notch expressed during mammalian
RT   development.";
RL   Development 113:199-205(1991).
RN   [2]
RP   REVISIONS TO 1652-1653.
RA   Weinmaster G.;
RL   Submitted (APR-1998) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   FUNCTION.
RX   MEDLINE=21094508; PubMed=11182080;
RA   Tanigaki K., Nogaki F., Takahashi J., Tashiro K., Kurooka H.,
RA   Honjo T.;
RT   "Notch1 and Notch3 instructively restrict bFGF-responsive multipotent
RT   neural progenitor cells to an astroglial fate.";
RL   Neuron 29:45-55(2001).
RN   [4]
RP   TISSUE SPECIFICITY.
RX   MEDLINE=93202015; PubMed=1295745;
RA   Weinmaster G., Roberts V.J., Lemke G.;
RT   "Notch2: a second mammalian Notch gene.";
RL   Development 116:931-941(1992).
RN   [5]
RP   TISSUE SPECIFICITY.
RX   MEDLINE=21331789; PubMed=11438922;
RA   Irvin D.K., Zurcher S.D., Nguyen T., Weinmaster G., Kornblum H.I.;
RT   "Expression patterns of Notch1, Notch2, and Notch3 suggest multiple
RT   functional roles for the Notch-DSL signaling system during brain
RT   development.";
RL   J. Comp. Neurol. 436:167-181(2001).
CC   -!- FUNCTION: Functions as a receptor for membrane-bound ligands
CC       Jagged1, Jagged2 and Delta1 to regulate cell-fate determination.
CC       Upon ligand activation through the released notch intracellular
CC       domain (NICD) it forms a transcriptional activator complex with
CC       RBP-J kappa and activates genes of the enhancer of split locus.
CC       Affects the implementation of differentiation, proliferation and
CC       apoptotic programs (By similarity). Acts instructively to control
CC       the cell fate determination of CNS multipotent progenitor cells,
CC       resulting in astroglial induction and neuron/oligodendrocyte
CC       suppression.
CC   -!- SUBUNIT: Heterodimer of a C-terminal fragment N(TM) and a N-
CC       terminal fragment N(EC) which are probably linked by disulfide
CC       bonds (By similarity).
CC   -!- SUBCELLULAR LOCATION: Type I membrane protein. Following
CC       proteolytical processing NICD is translocated to the nucleus (By
CC       similarity).
CC   -!- TISSUE SPECIFICITY: Expressed in the brain, kidney and spleen.
CC       Expressed in postnatal central nervous system (CNS) germinal zones
CC       and, in early postnatal life, within numerous cells throughout the
CC       CNS. Found in both subventricular and ventricular germinal zones.
CC   -!- DEVELOPMENTAL STAGE: In the embryo, highest levels occur between
CC       days 12 and 14 and decrease rapidly to much lower levels in the
CC       adult.
CC   -!- PTM: Synthesized in the endoplasmic reticulum as an inactive form
CC       which is proteolytically cleaved by a furin-like convertase in the
CC       trans-Golgi network before it reaches the plasma membrane to yield
CC       an active, ligand-accessible form. Cleavage results in a C-
CC       terminal fragment N(TM) and a N-terminal fragment N(EC). Following
CC       ligand binding, it is cleaved by TNF-alpha converting enzyme
CC       (TACE) to yield a membrane-associated intermediate fragment called
CC       notch extracellular truncation (NEXT). This fragment is then
CC       cleaved by presenilin dependent gamma-secretase to release a
CC       notch-derived peptide containing the intracellular domain (NICD)
CC       from the membrane (By similarity).
CC   -!- PTM: Phosphorylated (By similarity).
CC   -!- SIMILARITY: Belongs to the NOTCH family.
CC   -!- SIMILARITY: Contains 36 EGF-like domains.
CC   -!- SIMILARITY: Contains 3 Lin/Notch repeats.
CC   -!- SIMILARITY: Contains 5 ANK repeats.
CC
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; X57405; CAA40667.1; -.
DR   HSSP; P00740; 1EDM.
DR   InterPro; IPR002110; ANK.
```

```
DR   InterPro; IPR000152; Asx_hydroxyl_S.
DR   InterPro; IPR000742; EGF_2.
DR   InterPro; IPR001881; EGF_Ca.
DR   InterPro; IPR001438; EGF_II.
DR   InterPro; IPR006209; EGF_like.
DR   InterPro; IPR002049; Laminin_EGF.
DR   InterPro; IPR008297; Notch.
DR   InterPro; IPR000800; Notch_dom.
DR   Pfam; PF00023; ank; 6.
DR   Pfam; PF00008; EGF; 35.
DR   Pfam; PF00066; notch; 3.
DR   PIRSF; PIRSF002279; Notch; 1.
DR   PRINTS; PR00010; EGFBLOOD.
DR   PRINTS; PR00011; EGFLAMININ.
DR   PRINTS; PR01452; NOTCH.
DR   SMART; SM00248; ANK; 6.
DR   SMART; SM00179; EGF_CA; 25.
DR   SMART; SM00004; NL; 2.
DR   PROSITE; PS50297; ANK_REP_REGION; 1.
DR   PROSITE; PS50088; ANK_REPEAT; 4.
DR   PROSITE; PS00010; ASX_HYDROXYL; 22.
DR   PROSITE; PS00022; EGF_1; 35.
DR   PROSITE; PS01186; EGF_2; 26.
DR   PROSITE; PS50026; EGF_3; 36.
DR   PROSITE; PS01187; EGF_CA; 21.
KW   Receptor; Transcription regulation; Activator; Differentiation;
KW   Developmental protein; Repeat; ANK repeat; EGF-like domain;
KW   Transmembrane; Glycoprotein; Signal; Phosphorylation.
FT   SIGNAL        1     18       POTENTIAL.
FT   CHAIN        19   2531       NEUROGENIC LOCUS NOTCH HOMOLOG PROTEIN 1.
FT   CHAIN      1711   2531       NOTCH EXTRACELLULAR TRUNCATION (BY
FT                                SIMILARITY).
FT   CHAIN      1744   2531       NOTCH INTRACELLULAR DOMAIN (BY
FT                                SIMILARITY).
FT   DOMAIN       19   1723       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   1724   1746       POTENTIAL.
FT   DOMAIN     1747   2531       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN       20     58       EGF-LIKE 1.
FT   DOMAIN       59     99       EGF-LIKE 2.
FT   DOMAIN      102    139       EGF-LIKE 3.
FT   DOMAIN      140    176       EGF-LIKE 4.
FT   DOMAIN      178    216       EGF-LIKE 5, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      218    255       EGF-LIKE 6.
FT   DOMAIN      257    293       EGF-LIKE 7, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      295    333       EGF-LIKE 8, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      335    371       EGF-LIKE 9, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      372    410       EGF-LIKE 10.
FT   DOMAIN      412    450       EGF-LIKE 11, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      452    488       EGF-LIKE 12, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      490    526       EGF-LIKE 13, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      528    564       EGF-LIKE 14, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      566    601       EGF-LIKE 15, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      603    639       EGF-LIKE 16, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      641    676       EGF-LIKE 17, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      678    714       EGF-LIKE 18, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      716    751       EGF-LIKE 19, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      753    789       EGF-LIKE 20, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      791    827       EGF-LIKE 21, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      829    867       EGF-LIKE 22.
FT   DOMAIN      869    905       EGF-LIKE 23, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      907    943       EGF-LIKE 24.
FT   DOMAIN      945    981       EGF-LIKE 25, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      983   1019       EGF-LIKE 26.
FT   DOMAIN     1021   1057       EGF-LIKE 27, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     1059   1095       EGF-LIKE 28.
FT   DOMAIN     1097   1143       EGF-LIKE 29.
FT   DOMAIN     1145   1181       EGF-LIKE 30, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     1183   1219       EGF-LIKE 31, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     1221   1265       EGF-LIKE 32, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     1267   1305       EGF-LIKE 33.
FT   DOMAIN     1307   1346       EGF-LIKE 34.
FT   DOMAIN     1348   1384       EGF-LIKE 35.
FT   DOMAIN     1387   1426       EGF-LIKE 36.
```

```
FT   REPEAT     1445   1480       LIN/NOTCH 1.
FT   REPEAT     1481   1522       LIN/NOTCH 2.
FT   REPEAT     1523   1562       LIN/NOTCH 3.
FT   REPEAT     1917   1946       ANK 1.
FT   REPEAT     1950   1980       ANK 2.
FT   REPEAT     1984   2013       ANK 3.
FT   REPEAT     2017   2046       ANK 4.
FT   REPEAT     2050   2079       ANK 5.
FT   DOMAIN     1730   1733       POLY-ALA.
FT   DOMAIN     1891   1894       POLY-GLU.
FT   DOMAIN     2258   2261       POLY-PRO.
FT   DOMAIN     2497   2500       POLY-SER.
FT   SITE       1654   1655       CLEAVAGE BY (FURIN-LIKE PROTEASE) (BY
FT                                SIMILARITY).
FT   DISULFID     24     37       BY SIMILARITY.
FT   DISULFID     31     46       BY SIMILARITY.
FT   DISULFID     48     57       BY SIMILARITY.
FT   DISULFID     63     74       BY SIMILARITY.
FT   DISULFID     68     87       BY SIMILARITY.
FT   DISULFID     89     98       BY SIMILARITY.
FT   DISULFID    106    117       BY SIMILARITY.
FT   DISULFID    111    127       BY SIMILARITY.
FT   DISULFID    129    138       BY SIMILARITY.
FT   DISULFID    144    155       BY SIMILARITY.
FT   DISULFID    149    164       BY SIMILARITY.
FT   DISULFID    166    175       BY SIMILARITY.
FT   DISULFID    182    195       BY SIMILARITY.
FT   DISULFID    189    204       BY SIMILARITY.
FT   DISULFID    206    215       BY SIMILARITY.
FT   DISULFID    222    233       BY SIMILARITY.
FT   DISULFID    227    243       BY SIMILARITY.

  Query Match             3.2%;  Score 9;  DB 1;  Length 2531;
  Best Local Similarity   100.0%;  Pred. No. 1.9;
  Matches    9;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         45 PCLHGGTCQ 53
              |||||||||
Db       1029 PCLHGGTCQ 1037


RESULT 6
AMPA_BACSU
ID   AMPA_BACSU     STANDARD;      PRT;   500 AA.
AC   O32106;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Probable cytosol aminopeptidase (EC 3.4.11.1) (Leucine aminopeptidase)
DE   (LAP) (Leucyl aminopeptidase).
GN   PEPA OR BSU32050.
OS   Bacillus subtilis.
OC   Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX   NCBI_TaxID=1423;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=168;
RX   MEDLINE=98044033; PubMed=9384377;
RA   Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
RA   Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,
RA   Borriss R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,
RA   Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,
RA   Choi S.K., Codani J.J., Connerton I.F., Cummings N.J., Daniel R.A.,
RA   Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emmerson P.T.,
RA   Entian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,
RA   Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,
RA   Ghim S.Y., Glaser P., Goffeau A., Golightly E.J., Grandi G.,
RA   Guiseppi G., Guy B.J., Haga K., Haiech J., Harwood C.R., Henaut A.,
RA   Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,
RA   Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,
RA   Kobayashi Y., Koetter P., Koningstein G., Krogh S., Kumano M.,
RA   Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,
RA   Lee S.M., Levine A., Liu H., Masuda S., Mauel C., Medigue C.,
```

```
RA   Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,
RA   Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,
RA   Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,
RA   Presecan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
RA   Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadaie Y.,
RA   Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,
RA   Sekiguchi J., Sekowska A., Seror S.J., Serror P., Shin B.S., Soldo B.,
RA   Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,
RA   Takeuchi M., Tamakoshi A., Tanaka T., Terpstra P., Tognoni A.,
RA   Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,
RA   Viari A., Wambutt R., Wedler E., Wedler H., Weitzenegger T.,
RA   Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
RA   Yoshida K., Yoshikawa H.F., Zumstein E., Yoshikawa H., Danchin A.;
RT   "The complete genome sequence of the Gram-positive bacterium Bacillus
RT   subtilis.";
RL   Nature 390:249-256(1997).
CC   -!- FUNCTION: Presumably involved in the processing and regular
CC       turnover of intracellular proteins. Catalyzes the removal of
CC       unsubstituted N-terminal amino acids from various peptides (By
CC       similarity).
CC   -!- CATALYTIC ACTIVITY: Release of an N-terminal amino acid, Xaa-|-
CC       Xbb-, in which Xaa is preferably Leu, but may be other amino acids
CC       including Pro although not Arg or Lys, and Xbb may be Pro.
CC   -!- COFACTOR: Binds 2 manganese ions per subunit (By similarity).
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC   -!- SIMILARITY: Belongs to peptidase family M17.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; Z99120; CAB15195.1; -.
DR   PIR; F70012; F70012.
DR   HSSP; P00727; 1LAM.
DR   MEROPS; M17.UPW; -.
DR   SubtiList; BG13970; pepA.
DR   HAMAP; MF_00181; -; 1.
DR   InterPro; IPR000819; Peptidase_M17_C.
DR   InterPro; IPR008283; Peptidase_M17_N.
DR   Pfam; PF00883; Peptidase_M17; 1.
DR   Pfam; PF02789; Peptidase_M17_N; 1.
DR   PRINTS; PR00481; LAMNOPPTDASE.
DR   PROSITE; PS00631; CYTOSOL_AP; 1.
KW   Hydrolase; Aminopeptidase; Manganese; Complete proteome.
FT   METAL       261    261       MANGANESE 2 (BY SIMILARITY).
FT   METAL       266    266       MANGANESE 1 AND 2 (BY SIMILARITY).
FT   METAL       284    284       MANGANESE 2 (BY SIMILARITY).
FT   METAL       343    343       MANGANESE 1 (BY SIMILARITY).
FT   METAL       345    345       MANGANESE 1 AND 2 (BY SIMILARITY).
FT   ACT_SITE    273    273       POTENTIAL.
FT   ACT_SITE    347    347       POTENTIAL.
SQ   SEQUENCE   500 AA;  53657 MW;  3E82968F66566559 CRC64;

  Query Match             2.8%;  Score 8;  DB 1;  Length 500;
  Best Local Similarity   100.0%;  Pred. No. 4.6;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        232 DVATLTGG 239
              ||||||||
Db        367 DVATLTGG 374


RESULT 7
PGCN_RAT
ID   PGCN_RAT       STANDARD;      PRT;  1257 AA.
AC   P55067;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   15-MAR-2004 (Rel. 43, Last annotation update)
```

```
DE   Neurocan core protein precursor (Chondroitin sulfate proteoglycan 3)
DE   (245 kDa early postnatal core glycoprotein) [Contains: 150 kDa adult
DE   core glycoprotein].
GN   CSPG3 OR NCAN.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC   STRAIN=Sprague-Dawley; TISSUE=Brain;
RX   MEDLINE=92406907; PubMed=1326557;
RA   Rauch U., Karthikeyan L., Maurel P., Margolis R.U., Margolis R.K.;
RT   "Cloning and primary structure of neurocan, a developmentally
RT   regulated, aggregating chondroitin sulfate proteoglycan of brain.";
RL   J. Biol. Chem. 267:19536-19547(1992).
RN   [2]
RP   CHARACTERIZATION.
RX   MEDLINE=94230574; PubMed=7513709;
RA   Friedlander D.R., Milev P., Karthikeyan L., Margolis R.K.,
RA   Margolis R.U., Grumet M.;
RT   "The neuronal chondroitin sulfate proteoglycan neurocan binds to the
RT   neural cell adhesion molecules Ng-CAM/L1/NILE and N-CAM, and inhibits
RT   neuronal adhesion and neurite outgrowth.";
RL   J. Cell Biol. 125:669-680(1994).
CC   -!- FUNCTION: May modulate neuronal adhesion and neurite growth during
CC       development by binding to neural cell adhesion molecules (NG-CAM
CC       and N-CAM). Chondroitin sulfate proteoglycan; binds to hyaluronic
CC       acid.
CC   -!- TISSUE SPECIFICITY: Early postnatal and adult brain; not expressed
CC       in kidney, lung, liver and muscle.
CC   -!- PTM: CONTAINS MOSTLY CHONDROITIN SULFATE, BUT ALSO N-LINKED AND
CC       O-LINKED OLIGOSACCHARIDES (BY SIMILARITY).
CC   -!- PTM: TWO ISOFORMS WERE FOUND THAT PROBABLY ARISE BY PROTEOLYTIC
CC       DEGRADATION. THE LARGE ISOFORM IS PREDOMINANT IN EARLY POSTNATAL
CC       BRAIN, THE SMALL ISOFORM IS FOUND IN ADULT BRAIN.
CC   -!- SIMILARITY: Contains 1 immunoglobulin-like V-type domain.
CC   -!- SIMILARITY: Contains 2 EGF-like domains.
CC   -!- SIMILARITY: Contains 2 link domains.
CC   -!- SIMILARITY: Contains 1 C-type lectin family domain.
CC   -!- SIMILARITY: Contains 1 Sushi (SCR) domain.
CC   -!- SIMILARITY: Belongs to the aggrecan/versican proteoglycan family.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; M97161; AAC37679.1; -.
DR   PIR; S28764; S28764.
DR   HSSP; P00740; 1EDM.
DR   InterPro; IPR002353; AntifreezeII.
DR   InterPro; IPR000152; Asx_hydroxyl_S.
DR   InterPro; IPR000742; EGF_2.
DR   InterPro; IPR001881; EGF_Ca.
DR   InterPro; IPR006209; EGF_like.
DR   InterPro; IPR007110; Ig-like.
DR   InterPro; IPR003599; Ig.
DR   InterPro; IPR001304; Lectin_C.
DR   InterPro; IPR000538; Link.
DR   InterPro; IPR000436; Sushi_SCR_CCP.
DR   Pfam; PF00008; EGF; 2.
DR   Pfam; PF00047; ig; 1.
DR   Pfam; PF00059; lectin_c; 1.
DR   Pfam; PF00084; sushi; 1.
DR   Pfam; PF00193; Xlink; 2.
DR   PRINTS; PR00356; ANTIFREEZEII.
DR   PRINTS; PR01265; LINKMODULE.
DR   ProDom; PD000918; Link; 2.
DR   SMART; SM00032; CCP; 1.
```

```
DR   SMART; SM00034; CLECT; 1.
DR   SMART; SM00179; EGF_CA; 1.
DR   SMART; SM00409; IG; 1.
DR   SMART; SM00445; LINK; 2.
DR   PROSITE; PS00010; ASX_HYDROXYL; 1.
DR   PROSITE; PS00615; C_TYPE_LECTIN_1; 1.
DR   PROSITE; PS50041; C_TYPE_LECTIN_2; 1.
DR   PROSITE; PS00022; EGF_1; 3.
DR   PROSITE; PS01186; EGF_2; 1.
DR   PROSITE; PS50026; EGF_3; 2.
DR   PROSITE; PS01187; EGF_CA; 1.
DR   PROSITE; PS50835; IG_LIKE; 1.
DR   PROSITE; PS01241; LINK; 2.
KW   Glycoprotein; Hyaluronic acid; Proteoglycan; Immunoglobulin domain;
KW   EGF-like domain; Calcium; Repeat; Lectin; Sushi; Signal.
FT   SIGNAL        1     22
FT   CHAIN        23   1257       NEUROCAN CORE PROTEIN.
FT   CHAIN       639   1257       150 kDa ADULT CORE GLYCOPROTEIN.
FT   DOMAIN       37    157       IG-LIKE V-TYPE.
FT   DOMAIN      158    253       LINK 1.
FT   DOMAIN      259    355       LINK 2.
FT   DOMAIN      949    985       EGF-LIKE 1.
FT   DOMAIN      987   1023       EGF-LIKE 2, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     1025   1154       C-TYPE LECTIN.
FT   DOMAIN     1155   1213       SUSHI.
FT   DISULFID     58    139       BY SIMILARITY.
FT   DISULFID    181    252       BY SIMILARITY.
FT   DISULFID    205    226       BY SIMILARITY.
FT   DISULFID    279    354       BY SIMILARITY.
FT   DISULFID    303    324       BY SIMILARITY.
FT   DISULFID    953    964       BY SIMILARITY.
FT   DISULFID    958    973       BY SIMILARITY.
FT   DISULFID    975    984       BY SIMILARITY.
FT   DISULFID   1029   1040       BY SIMILARITY.
FT   DISULFID   1057   1149       BY SIMILARITY.
FT   DISULFID   1125   1141       BY SIMILARITY.
FT   DISULFID   1156   1199       BY SIMILARITY.
FT   DISULFID   1185   1212       BY SIMILARITY.
FT   CARBOHYD    121    121       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    339    339       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    737    737       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    944    944       O-LINKED (XYL. . .) (CHONDROITIN SULFATE).
FT   CARBOHYD    967    967       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1164   1164       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   1257 AA;  135544 MW;  992B33DCFA19EE1B CRC64;

  Query Match             2.8%;  Score 8;  DB 1;  Length 1257;
  Best Local Similarity   100.0%;  Pred. No. 11;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          45 PCLHGGTC 52
               ||||||||
Db         957 PCLHGGTC 964


RESULT 8
PGCN_MOUSE
ID   PGCN_MOUSE     STANDARD;      PRT;  1268 AA.
AC   P55066;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Neurocan core protein precursor (Chondroitin sulfate proteoglycan 3).
GN   CSPG3 OR NCAN.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BALB/c; TISSUE=Brain;
RX   MEDLINE=96039250; PubMed=7490074;
RA   Rauch U., Grimpe B., Kulbe G., Arnold-Ammer I., Beier D.,
RA   Faessler R.;
RT   "Structure and chromosomal localization of the mouse neurocan gene.";
RL   Genomics 28:405-410(1995).
CC   -!- FUNCTION: May modulate neuronal adhesion and neurite growth during
CC       development by binding to neural cell adhesion molecules (NG-CAM
CC       and N-CAM). Chondroitin sulfate proteoglycan; binds to hyaluronic
CC       acid.
CC   -!- TISSUE SPECIFICITY: Brain.
CC   -!- SIMILARITY: Contains 1 immunoglobulin-like V-type domain.
CC   -!- SIMILARITY: Contains 2 EGF-like domains.
CC   -!- SIMILARITY: Contains 2 link domains.
CC   -!- SIMILARITY: Contains 1 C-type lectin family domain.
CC   -!- SIMILARITY: Contains 1 Sushi (SCR) domain.
CC   -!- SIMILARITY: Belongs to the aggrecan/versican proteoglycan family.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; X84727; CAA59216.1; -.
DR   PIR; S52781; S52781.
DR   HSSP; P00740; 1EDM.
DR   MGD; MGI:104694; Cspg3.
DR   InterPro; IPR002353; AntifreezeII.
DR   InterPro; IPR000152; Asx_hydroxyl_S.
DR   InterPro; IPR000742; EGF_2.
DR   InterPro; IPR001881; EGF_Ca.
DR   InterPro; IPR006209; EGF_like.
DR   InterPro; IPR007110; Ig-like.
DR   InterPro; IPR003599; Ig.
DR   InterPro; IPR001304; Lectin_C.
DR   InterPro; IPR000538; Link.
DR   InterPro; IPR000436; Sushi_SCR_CCP.
DR   Pfam; PF00008; EGF; 2.
DR   Pfam; PF00047; ig; 1.
DR   Pfam; PF00059; lectin_c; 1.
DR   Pfam; PF00084; sushi; 1.
DR   Pfam; PF00193; Xlink; 2.
DR   PRINTS; PR00356; ANTIFREEZEII.
DR   PRINTS; PR01265; LINKMODULE.
DR   ProDom; PD000918; Link; 2.
DR   SMART; SM00032; CCP; 1.
DR   SMART; SM00034; CLECT; 1.
DR   SMART; SM00179; EGF_CA; 1.
DR   SMART; SM00409; IG; 1.
DR   SMART; SM00445; LINK; 2.
DR   PROSITE; PS00010; ASX_HYDROXYL; 1.
DR   PROSITE; PS00615; C_TYPE_LECTIN_1; 1.
DR   PROSITE; PS50041; C_TYPE_LECTIN_2; 1.
DR   PROSITE; PS00022; EGF_1; 3.
DR   PROSITE; PS01186; EGF_2; 1.
DR   PROSITE; PS50026; EGF_3; 2.
DR   PROSITE; PS01187; EGF_CA; 1.
DR   PROSITE; PS50835; IG_LIKE; 1.
DR   PROSITE; PS01241; LINK; 2.
KW   Glycoprotein; Hyaluronic acid; Proteoglycan; Immunoglobulin domain;
KW   EGF-like domain; Calcium; Repeat; Lectin; Sushi; Signal.
FT   SIGNAL        1     22       POTENTIAL.
FT   CHAIN        23   1268       NEUROCAN CORE PROTEIN.
FT   DOMAIN       37    157       IG-LIKE V-TYPE.
FT   DOMAIN      158    253       LINK 1.
FT   DOMAIN      259    355       LINK 2.
FT   DOMAIN      960    996       EGF-LIKE 1.
FT   DOMAIN      998   1034       EGF-LIKE 2, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     1036   1165       C-TYPE LECTIN.
FT   DOMAIN     1166   1224       SUSHI.
FT   DISULFID     58    139       BY SIMILARITY.
FT   DISULFID    181    252       BY SIMILARITY.
FT   DISULFID    205    226       BY SIMILARITY.
```

```
FT   DISULFID    279    354       BY SIMILARITY.
FT   DISULFID    303    324       BY SIMILARITY.
FT   DISULFID    964    975       BY SIMILARITY.
FT   DISULFID    969    984       BY SIMILARITY.
FT   DISULFID    986    995       BY SIMILARITY.
FT   DISULFID   1040   1051       BY SIMILARITY.
FT   DISULFID   1068   1160       BY SIMILARITY.
FT   DISULFID   1136   1152       BY SIMILARITY.
FT   DISULFID   1167   1210       BY SIMILARITY.
FT   DISULFID   1196   1223       BY SIMILARITY.
FT   CARBOHYD    121    121       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    339    339       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    742    742       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    978    978       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1175   1175       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   1268 AA;  137200 MW;  3014E8E202A2FAEC CRC64;

  Query Match             2.8%;  Score 8;  DB 1;  Length 1268;
  Best Local Similarity   100.0%;  Pred. No. 11;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         45 PCLHGGTC 52
              ||||||||
Db        968 PCLHGGTC 975


RESULT 9
PGCN_HUMAN
ID   PGCN_HUMAN     STANDARD;      PRT;  1321 AA.
AC   O14594; Q9UPK6;
DT   28-FEB-2003 (Rel. 41, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Neurocan core protein precursor (Chondroitin sulfate proteoglycan 3).
GN   CSPG3 OR NCAN OR NEUR.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=99013874; PubMed=9795216;
RA   Prange C.K., Pennacchio L.A., Lieuallen K., Fan W., Lennon G.G.;
RT   "Characterization of the human neurocan gene, CSPG3.";
RL   Gene 221:199-205(1998).
RN   [2]
RP   SEQUENCE OF 1-990 AND 1007-1321 FROM N.A., AND VARIANT ALA-1254.
RA   Lamerdin J.E., McCready P.M., Skowronski E., Adamson A.W.,
RA   Burkhart-Schultz K., Gordon L., Kyle A., Ramirez M., Stilwagen S.,
RA   Phan H., Velasco N., Do L., Regala W., Terry A., Garnes J.,
RA   Danganan L., Poundstone P., Christensen M., Georgescu A., Avila J.,
RA   Liu S., Attix C., Andreise T., Trankheim M., Amico-Keller G.,
RA   Coefield J., Duarte S., Lucas S., Bruce R., Thomas P., Quan G.,
RA   Kronmiller B., Arellano A., Montgomery M., Ow D., Nolan M., Trong S.,
RA   Kobayashi A., Olsen A.S., Carrano A.V.;
RT   "Sequence analysis of an ~1 Mb region containing the MEF2B gene in
RT   19p12.";
RL   Submitted (JUL-1998) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: May modulate neuronal adhesion and neurite growth during
CC       development by binding to neural cell adhesion molecules (NG-CAM
CC       and N-CAM). Chondroitin sulfate proteoglycan; binds to hyaluronic
CC       acid.
CC   -!- TISSUE SPECIFICITY: Brain.
CC   -!- SIMILARITY: Contains 1 immunoglobulin-like V-type domain.
CC   -!- SIMILARITY: Contains 2 EGF-like domains.
CC   -!- SIMILARITY: Contains 2 link domains.
CC   -!- SIMILARITY: Contains 1 C-type lectin family domain.
CC   -!- SIMILARITY: Contains 1 Sushi (SCR) domain.
CC   -!- SIMILARITY: Belongs to the aggrecan/versican proteoglycan family.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
```

```
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; AF026547; AAC80576.1; -.
DR   EMBL; AC003110; AAB86655.1; -.
DR   EMBL; AC005254; AAC25581.1; -.
DR   HSSP; P00740; 1EDM.
DR   Genew; HGNC:2465; CSPG3.
DR   MIM; 600826; -.
DR   InterPro; IPR000152; Asx_hydroxyl_S.
DR   InterPro; IPR000742; EGF_2.
DR   InterPro; IPR001881; EGF_Ca.
DR   InterPro; IPR006209; EGF_like.
DR   InterPro; IPR007110; Ig-like.
DR   InterPro; IPR003599; Ig.
DR   InterPro; IPR001304; Lectin_C.
DR   InterPro; IPR000538; Link.
DR   InterPro; IPR000436; Sushi_SCR_CCP.
DR   Pfam; PF00008; EGF; 2.
DR   Pfam; PF00047; ig; 1.
DR   Pfam; PF00059; lectin_c; 1.
DR   Pfam; PF00084; sushi; 1.
DR   Pfam; PF00193; Xlink; 2.
DR   PRINTS; PR01265; LINKMODULE.
DR   ProDom; PD000918; Link; 2.
DR   SMART; SM00032; CCP; 1.
DR   SMART; SM00034; CLECT; 1.
DR   SMART; SM00179; EGF_CA; 1.
DR   SMART; SM00409; IG; 1.
DR   SMART; SM00445; LINK; 2.
DR   PROSITE; PS00010; ASX_HYDROXYL; 1.
DR   PROSITE; PS00615; C_TYPE_LECTIN_1; 1.
DR   PROSITE; PS50041; C_TYPE_LECTIN_2; 1.
DR   PROSITE; PS00022; EGF_1; 3.
DR   PROSITE; PS01186; EGF_2; 1.
DR   PROSITE; PS50026; EGF_3; 2.
DR   PROSITE; PS01187; EGF_CA; 1.
DR   PROSITE; PS50835; IG_LIKE; 1.
DR   PROSITE; PS01241; LINK; 2.
KW   Glycoprotein; Hyaluronic acid; Proteoglycan; Immunoglobulin domain;
KW   EGF-like domain; Calcium; Repeat; Lectin; Sushi; Signal; Polymorphism.
FT   SIGNAL        1     22       POTENTIAL.
FT   CHAIN        23   1321       NEUROCAN CORE PROTEIN.
FT   DOMAIN       38    153       IG-LIKE V-TYPE.
FT   DOMAIN      159    254       LINK 1.
FT   DOMAIN      260    356       LINK 2.
FT   DOMAIN     1008   1044       EGF-LIKE 1.
FT   DOMAIN     1046   1082       EGF-LIKE 2, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     1084   1213       C-TYPE LECTIN.
FT   DOMAIN     1214   1272       SUSHI.
FT   DISULFID     59    140       BY SIMILARITY.
FT   DISULFID    182    253       BY SIMILARITY.
FT   DISULFID    206    227       BY SIMILARITY.
FT   DISULFID    280    355       BY SIMILARITY.
FT   DISULFID    304    325       BY SIMILARITY.
FT   DISULFID   1012   1023       BY SIMILARITY.
FT   DISULFID   1017   1032       BY SIMILARITY.
FT   DISULFID   1034   1043       BY SIMILARITY.
FT   DISULFID   1088   1099       BY SIMILARITY.
FT   DISULFID   1116   1208       BY SIMILARITY.
FT   DISULFID   1184   1200       BY SIMILARITY.
FT   DISULFID   1215   1258       BY SIMILARITY.
FT   DISULFID   1244   1271       BY SIMILARITY.
FT   CARBOHYD    122    122       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    340    340       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1026   1026       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1223   1223       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   VARIANT    1254   1254       V -> A (in dbSNP:1064389).
FT                                /FTId=VAR_016176.
FT   CONFLICT   1234   1234       N -> Y (IN REF. 2).
FT   CONFLICT   1282   1282       G -> R (IN REF. 2).
```

```
SQ   SEQUENCE   1321 AA;  142972 MW;  2EF47F823DB980B8 CRC64;

  Query Match             2.8%;  Score 8;  DB 1;  Length 1321;
  Best Local Similarity   100.0%;  Pred. No. 11;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         45 PCLHGGTC 52
              ||||||||
Db       1016 PCLHGGTC 1023


RESULT 10
AGRI_DISOM
ID   AGRI_DISOM     STANDARD;      PRT;  1328 AA.
AC   Q90404;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Agrin (Fragment).
GN   AGRN.
OS   Discopyge ommata (Electric ray).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Chondrichthyes;
OC   Elasmobranchii; Squalea; Hypnosqualea; Pristiorajea; Batoidea;
OC   Torpediniformes; Narcinoidei; Narcinidae; Discopyge.
OX   NCBI_TaxID=7785;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Smith M.A., Magill-Solc C., Rupp F., Yao Y.-M.M., Schilling J.W.,
RA   Snow P., McMahan U.J.;
RT   "Isolation and characterization of a cDNA that encodes an agrin
RT   homolog in the marine ray.";
RL   Mol. Cell. Neurosci. 3:406-417(1992).
CC   -!- FUNCTION: Component of the basal lamina that causes the
CC       aggregation of acetylcholine receptors and acetylcholine-esterase
CC       on the surface of muscle fibers of the neuromuscular junction (By
CC       similarity).
CC   -!- SUBCELLULAR LOCATION: Synaptic basal lamina at the neuromuscular
CC       junction (By similarity).
CC   -!- SIMILARITY: Contains at least 2 laminin EGF-like domains.
CC   -!- SIMILARITY: Contains at least 3 EGF-like domains.
CC   -!- SIMILARITY: Contains 1 SEA domain.
CC   -!- SIMILARITY: Contains 3 laminin G-like domains.
CC
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; L01423; AAA49224.1; -.
DR   PIR; T43060; T43060.
DR   HSSP; P00740; 1EDM.
DR   InterPro; IPR000152; Asx_hydroxyl_S.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR000742; EGF_2.
DR   InterPro; IPR006209; EGF_like.
DR   InterPro; IPR003645; FolN.
DR   InterPro; IPR002350; kazal.
DR   InterPro; IPR002049; Laminin_EGF.
DR   InterPro; IPR001791; Laminin_G.
DR   InterPro; IPR000082; SEA_domain.
DR   Pfam; PF00008; EGF; 4.
DR   Pfam; PF00050; kazal; 2.
DR   Pfam; PF00053; laminin_EGF; 2.
DR   Pfam; PF00054; laminin_G; 3.
DR   Pfam; PF01390; SEA; 1.
DR   PRINTS; PR00011; EGFLAMININ.
DR   SMART; SM00180; EGF_Lam; 2.
DR   SMART; SM00274; FOLN; 2.
DR   SMART; SM00280; KAZAL; 2.
DR   SMART; SM00282; LamG; 3.
```

```
DR   SMART; SM00200; SEA; 1.
DR   PROSITE; PS00010; ASX_HYDROXYL; 1.
DR   PROSITE; PS00022; EGF_1; 5.
DR   PROSITE; PS01186; EGF_2; 1.
DR   PROSITE; PS50026; EGF_3; 4.
DR   PROSITE; PS50025; LAM_G_DOMAIN; 3.
DR   PROSITE; PS01248; LAMININ_TYPE_EGF; 1.
DR   PROSITE; PS50024; SEA; 1.
KW   Glycoprotein; EGF-like domain; Repeat; Laminin EGF-like domain.
FT   NON_TER       1      1
FT   DOMAIN       79    132       LAMININ EGF-LIKE 1.
FT   DOMAIN      133    179       LAMININ EGF-LIKE 2.
FT   DOMAIN      411    533       SEA.
FT   DOMAIN      608    644       EGF-LIKE 1.
FT   DOMAIN      649    825       LAMININ G-LIKE 1.
FT   DOMAIN      865    902       EGF-LIKE 2.
FT   DOMAIN      914   1096       LAMININ G-LIKE 2.
FT   DOMAIN     1097   1135       EGF-LIKE 3.
FT   DOMAIN     1146   1324       LAMININ G-LIKE 3.
FT   DISULFID     79     91       BY SIMILARITY.
FT   DISULFID     81     98       BY SIMILARITY.
FT   DISULFID    100    109       BY SIMILARITY.
FT   DISULFID    112    130       BY SIMILARITY.
FT   DISULFID    133    145       BY SIMILARITY.
FT   DISULFID    135    152       BY SIMILARITY.
FT   DISULFID    154    163       BY SIMILARITY.
FT   DISULFID    166    177       BY SIMILARITY.
FT   DISULFID    612    623       BY SIMILARITY.
FT   DISULFID    617    632       BY SIMILARITY.
FT   DISULFID    634    643       BY SIMILARITY.
FT   DISULFID    869    880       BY SIMILARITY.
FT   DISULFID    874    890       BY SIMILARITY.
FT   DISULFID    892    901       BY SIMILARITY.
FT   DISULFID   1101   1114       BY SIMILARITY.
FT   DISULFID   1108   1123       BY SIMILARITY.
FT   DISULFID   1125   1134       BY SIMILARITY.
FT   CARBOHYD     63     63       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    168    168       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    279    279       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    301    301       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    369    369       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   1328 AA;  144018 MW;  79D81C1AF2A71C18 CRC64;

  Query Match             2.8%;  Score 8;  DB 1;  Length 1328;
  Best Local Similarity   100.0%;  Pred. No. 11;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         45 PCLHGGTC 52
              ||||||||
Db        616 PCLHGGTC 623


RESULT 11
AGRI_CHICK
ID   AGRI_CHICK     STANDARD;      PRT;  1955 AA.
AC   P31696;
DT   01-JUL-1993 (Rel. 26, Created)
DT   01-JUL-1993 (Rel. 26, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Agrin precursor.
GN   AGRN.
OS   Gallus gallus (Chicken).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC   Gallus.
OX   NCBI_TaxID=9031;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Brain;
RX   MEDLINE=92232297; PubMed=1314620;
RA   Tsim K.W.K., Ruegg M.A., Escher G., Kroeger S., McMahan U.J.;
RT   "cDNA that encodes active agrin.";
RL   Neuron 8:677-689(1992).
```

```
RN   [2]
RP   ALTERNATIVE SPLICING.
RX   MEDLINE=92232298; PubMed=1314621;
RA   Ruegg M.A., Tsim K.W.K., Horton S.E., Kroeger S., Escher G.,
RA   Gensch E.M., McMahan U.J.;
RT   "The agrin gene codes for a family of basal lamina proteins that
RT   differ in function and distribution.";
RL   Neuron 8:691-699(1992).
CC   -!- FUNCTION: Component of the basal lamina that causes the
CC       aggregation of acetylcholine receptors and acetylcholine-esterase
CC       on the surface of muscle fibers of the neuromuscular junction.
CC   -!- SUBUNIT: Binds to laminin.
CC   -!- SUBCELLULAR LOCATION: Synaptic basal lamina at the neuromuscular
CC       junction.
CC   -!- ALTERNATIVE PRODUCTS:
CC       Event=Alternative splicing; Named isoforms=3;
CC         Comment=Additional isoforms seem to exist. Isoforms differ in
CC         their acetylcholine receptor clustering activity;
CC       Name=1;
CC         IsoId=P31696-1; Sequence=Displayed;
CC       Name=2; Synonyms=Agrin-related protein 1;
CC         IsoId=P31696-2; Sequence=VSP_001370;
CC       Name=3; Synonyms=Agrin-related protein 2;
CC         IsoId=P31696-3; Sequence=VSP_001369, VSP_001370;
CC   -!- PTM: Contains heparan sulfate chains as well as N-linked and O-
CC       linked oligosaccharides (By similarity).
CC   -!- SIMILARITY: Contains 9 Kazal-like domains.
CC   -!- SIMILARITY: Contains 2 laminin EGF-like domains.
CC   -!- SIMILARITY: Contains 4 EGF-like domains.
CC   -!- SIMILARITY: Contains 1 SEA domain.
CC   -!- SIMILARITY: Contains 3 laminin G-like domains.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; M94271; AAA48585.1; -.
DR   EMBL; M97371; AAA48586.1; -.
DR   EMBL; M97372; -; NOT_ANNOTATED_CDS.
DR   PIR; JH0591; AGCH.
DR   HSSP; P00740; 1EDM.
DR   InterPro; IPR004850; Agrin_NtA.
DR   InterPro; IPR000152; Asx_hydroxyl_S.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR000742; EGF_2.
DR   InterPro; IPR006209; EGF_like.
DR   InterPro; IPR003645; FolN.
DR   InterPro; IPR002350; kazal.
DR   InterPro; IPR002049; Laminin_EGF.
DR   InterPro; IPR001791; Laminin_G.
DR   InterPro; IPR000082; SEA_domain.
DR   InterPro; IPR008993; TIMP_like.
DR   Pfam; PF00008; EGF; 4.
DR   Pfam; PF00050; kazal; 9.
DR   Pfam; PF00053; laminin_EGF; 2.
DR   Pfam; PF00054; laminin_G; 3.
DR   Pfam; PF03146; NtA; 1.
DR   Pfam; PF01390; SEA; 1.
DR   PRINTS; PR00011; EGFLAMININ.
DR   SMART; SM00180; EGF_Lam; 2.
DR   SMART; SM00274; FOLN; 8.
DR   SMART; SM00280; KAZAL; 9.
DR   SMART; SM00282; LamG; 3.
DR   SMART; SM00200; SEA; 1.
DR   PROSITE; PS00010; ASX_HYDROXYL; 1.
DR   PROSITE; PS00022; EGF_1; 6.
DR   PROSITE; PS01186; EGF_2; 1.
DR   PROSITE; PS50026; EGF_3; 4.
DR   PROSITE; PS50025; LAM_G_DOMAIN; 3.
```

```
DR   PROSITE; PS01248; LAMININ_TYPE_EGF; 1.
DR   PROSITE; PS50024; SEA; 1.
KW   Glycoprotein; EGF-like domain; Repeat; Alternative splicing; Signal;
KW   Laminin EGF-like domain; Proteoglycan; Heparan sulfate.
FT   SIGNAL        1     38       POTENTIAL.
FT   CHAIN        39   1955       AGRIN.
FT   DOMAIN       54    126       KAZAL-LIKE 1.
FT   DOMAIN      130    201       KAZAL-LIKE 2.
FT   DOMAIN      202    273       KAZAL-LIKE 3.
FT   DOMAIN      276    344       KAZAL-LIKE 4.
FT   DOMAIN      350    418       KAZAL-LIKE 5.
FT   DOMAIN      419    483       KAZAL-LIKE 6.
FT   DOMAIN      484    548       KAZAL-LIKE 7.
FT   DOMAIN      551    633       KAZAL-LIKE 8.
FT   DOMAIN      675    728       LAMININ EGF-LIKE 1.
FT   DOMAIN      729    775       LAMININ EGF-LIKE 2.
FT   DOMAIN      781    851       KAZAL-LIKE 9.
FT   DOMAIN     1026   1148       SEA.
FT   DOMAIN     1229   1265       EGF-LIKE 1.
FT   DOMAIN     1270   1445       LAMININ G-LIKE 1.
FT   DOMAIN     1446   1483       EGF-LIKE 2.
FT   DOMAIN     1485   1522       EGF-LIKE 3.
FT   DOMAIN     1532   1713       LAMININ G-LIKE 2.
FT   DOMAIN     1714   1752       EGF-LIKE 4.
FT   DOMAIN     1776   1952       LAMININ G-LIKE 3.
FT   DOMAIN      856    995       SER/THR-RICH.
FT   DOMAIN     1150   1219       SER/THR-RICH.
FT   DISULFID     86    105       POTENTIAL.
FT   DISULFID     94    126       POTENTIAL.
FT   DISULFID    160    180       POTENTIAL.
FT   DISULFID    169    201       POTENTIAL.
FT   DISULFID    233    252       POTENTIAL.
FT   DISULFID    241    273       POTENTIAL.
FT   DISULFID    304    323       POTENTIAL.
FT   DISULFID    312    344       POTENTIAL.
FT   DISULFID    378    397       POTENTIAL.
FT   DISULFID    386    418       POTENTIAL.
FT   DISULFID    443    462       POTENTIAL.
FT   DISULFID    451    483       POTENTIAL.
FT   DISULFID    507    527       POTENTIAL.
FT   DISULFID    516    548       POTENTIAL.
FT   DISULFID    592    612       POTENTIAL.
FT   DISULFID    601    633       POTENTIAL.
FT   DISULFID    675    687       BY SIMILARITY.
FT   DISULFID    677    694       BY SIMILARITY.
FT   DISULFID    696    705       BY SIMILARITY.
FT   DISULFID    708    726       BY SIMILARITY.
FT   DISULFID    729    741       BY SIMILARITY.
FT   DISULFID    731    748       BY SIMILARITY.
FT   DISULFID    750    759       BY SIMILARITY.
FT   DISULFID    762    773       BY SIMILARITY.
FT   DISULFID    810    830       POTENTIAL.
FT   DISULFID    819    851       POTENTIAL.
FT   DISULFID   1233   1244       BY SIMILARITY.
FT   DISULFID   1238   1253       BY SIMILARITY.
FT   DISULFID   1255   1264       BY SIMILARITY.
FT   DISULFID   1450   1461       BY SIMILARITY.
FT   DISULFID   1455   1471       BY SIMILARITY.
FT   DISULFID   1473   1482       BY SIMILARITY.
FT   DISULFID   1489   1500       BY SIMILARITY.
FT   DISULFID   1494   1510       BY SIMILARITY.
FT   DISULFID   1512   1521       BY SIMILARITY.
FT   DISULFID   1718   1731       BY SIMILARITY.
FT   DISULFID   1725   1740       BY SIMILARITY.
FT   DISULFID   1742   1751       BY SIMILARITY.
FT   CARBOHYD    390    390       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    659    659       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    764    764       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    814    814       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   VARSPLIC   1648   1651       Missing (in isoform 3).
FT                                /FTId=VSP_001369.
FT   VARSPLIC   1783   1793       Missing (in isoform 2 and isoform 3).
FT                                /FTId=VSP_001370.
```

```
FT   CONFLICT   1129   1131       RTI -> SIL (IN REF. 1; AAA48586).
SQ   SEQUENCE   1955 AA;  211411 MW;  B4DEB27C23422581 CRC64;

  Query Match             2.8%;  Score 8;  DB 1;  Length 1955;
  Best Local Similarity   100.0%;  Pred. No. 16;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         45 PCLHGGTC 52
              ||||||||
Db       1237 PCLHGGTC 1244


RESULT 12
NTC3_HUMAN
ID   NTC3_HUMAN     STANDARD;      PRT;  2321 AA.
AC   Q9UM47; Q9UEB3; Q9UPL3; Q9Y6L8;
DT   28-FEB-2003 (Rel. 41, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Neurogenic locus notch homolog protein 3 precursor (Notch 3).
GN   NOTCH3.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=97032728; PubMed=8878478;
RA   Joutel A., Corpechot C., Ducros A., Vahedi K., Chabriat H., Mouton P.,
RA   Alamowitch S., Domenga V., Cecillion M., Marechal E., Maciazek J.,
RA   Vayssiere C., Cruaud C., Cabanis E.-A., Ruchoux M.M., Weissenbach J.,
RA   Bach J.-F., Bousser M.-G., Tournier-Lasserve E.;
RT   "Notch3 mutations in CADASIL, a hereditary adult-onset condition
RT   causing stroke and dementia.";
RL   Nature 383:707-710(1996).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Gunel M., Artavanis-Tsakonas S.;
RL   Submitted (APR-1998) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RA   Lamerdin J.E., McCready P.M., Skowronski E., Adamson A.W.,
RA   Burkhart-Schultz K., Gordon L., Kyle A., Ramirez M., Stilwagen S.,
RA   Phan H., Velasco N., Garnes J., Danganan L., Poundstone P.,
RA   Christensen M., Georgescu A., Avila J., Liu S., Attix C., Andreise T.,
RA   Trankheim M., Amico-Keller G., Coefield J., Duarte S., Lucas S.,
RA   Bruce R., Thomas P., Quan G., Kronmiller B., Arellano A.,
RA   Montgomery M., Ow D., Nolan M., Trong S., Kobayashi A., Olsen A.S.,
RA   Carrano A.V.;
RT   "Sequence analysis of an 1.5 Mb olfactory receptor (OLFR) cluster in
RT   19p13.1.";
RL   Submitted (MAY-1998) to the EMBL/GenBank/DDBJ databases.
RN   [4]
RP   VARIANTS CADASIL TYR-49; CYS-71; CYS-90; CYS-110; CYS-133; CYS-141;
RP   ARG-146; CYS-153; CYS-169; CYS-171; CYS-182; ARG-185; SER-212;
RP   GLY-222; TYR-224; CYS-258; TYR-542; CYS-558; CYS-578; CYS-728;
RP   CYS-985; CYS-1006; CYS-1031; CYS-1231 AND ARG-1261, AND VARIANTS
RP   ARG-170; LEU-496; GLN-1133; MET-1183 AND ALA-2223.
RX   MEDLINE=98049753; PubMed=9388399;
RA   Joutel A., Vahedi K., Corpechot C., Troesch A., Chabriat H.,
RA   Vayssiere C., Cruaud C., Maciazek J., Weissenbach J., Bousser M.-G.,
RA   Bach J.-F., Tournier-Lasserve E.;
RT   "Strong clustering and stereotyped nature of Notch3 mutations in
RT   CADASIL patients.";
RL   Lancet 350:1511-1515(1997).
RN   [5]
RP   VARIANT CADASIL 114-GLY--PRO-120 DEL.
RX   MEDLINE=20264473; PubMed=10802807;
RA   Joutel A., Chabriat H., Vahedi K., Domenga V., Vayssiere C.,
RA   Ruchoux M.M., Lucas C., Leys D., Bousser M.-G., Tournier-Lasserve E.;
RT   "Splice site mutation causing a seven amino acid Notch3 in-frame
RT   deletion in CADASIL.";
RL   Neurology 54:1874-1875(2000).
RN   [6]
RP   IDENTIFICATION OF LIGANDS.
RX   MEDLINE=99180765; PubMed=10079256;
RA   Gray G.E., Mann R.S., Mitsiadis E., Henrique D., Carcangiu M.-L.,
RA   Banks A., Leiman J., Ward D., Ish-Horowitz D., Artavanis-Tsakonas S.;
RT   "Human ligands of the Notch receptor.";
RL   Am. J. Pathol. 154:785-794(1999).
CC   -!- FUNCTION: Functions as a receptor for membrane-bound ligands
CC       Jagged1, Jagged2 and Delta1 to regulate cell-fate determination.
CC       Upon ligand activation through the released notch intracellular
CC       domain (NICD) it forms a transcriptional activator complex with
CC       RBP-J kappa and activates genes of the enhancer of split locus.
CC       Affects the implementation of differentiation, proliferation and
CC       apoptotic programs (By similarity).
CC   -!- SUBUNIT: Heterodimer of a C-terminal fragment N(TM) and a N-
CC       terminal fragment N(EC) which are probably linked by disulfide
CC       bonds (By similarity).
CC   -!- SUBCELLULAR LOCATION: Type I membrane protein. Following
CC       proteolytical processing NICD is translocated to the nucleus.
CC   -!- TISSUE SPECIFICITY: Ubiquitously expressed in fetal and adult
CC       tissues.
CC   -!- PTM: Synthesized in the endoplasmic reticulum as an inactive form
CC       which is proteolytically cleaved by a furin-like convertase in the
CC       trans-Golgi network before it reaches the plasma membrane to yield
CC       an active, ligand-accessible form. Cleavage results in a C-
CC       terminal fragment N(TM) and a N-terminal fragment N(EC). Following
CC       ligand binding, it is cleaved by TNF-alpha converting enzyme
CC       (TACE) to yield a membrane-associated intermediate fragment called
CC       notch extracellular truncation (NEXT). This fragment is then
CC       cleaved by presenilin dependent gamma-secretase to release a
CC       notch-derived peptide containing the intracellular domain (NICD)
CC       from the membrane (By similarity).
CC   -!- PTM: Phosphorylated (By similarity).
CC   -!- DISEASE: Defects in NOTCH3 are associated with cerebral autosomal
CC       dominant arteriopathy with subcortical infarcts and
CC       leukoencephalopathy (CADASIL) [MIM:125310]. CADASIL causes a type
CC       of stroke and dementia of which key features include recurrent
CC       subcortical ischemic events and vascular dementia.
CC   -!- SIMILARITY: Belongs to the NOTCH family.
CC   -!- SIMILARITY: Contains 34 EGF-like domains.
CC   -!- SIMILARITY: Contains 3 Lin/Notch repeats.
CC   -!- SIMILARITY: Contains 5 ANK repeats.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; U97669; AAB91371.1; -.
DR   EMBL; AF058900; AAC14346.1; -.
DR   EMBL; AF058881; AAC14346.1; JOINED.
DR   EMBL; AF058882; AAC14346.1; JOINED.
DR   EMBL; AF058883; AAC14346.1; JOINED.
DR   EMBL; AF058884; AAC14346.1; JOINED.
DR   EMBL; AF058885; AAC14346.1; JOINED.
DR   EMBL; AF058886; AAC14346.1; JOINED.
DR   EMBL; AF058887; AAC14346.1; JOINED.
DR   EMBL; AF058888; AAC14346.1; JOINED.
DR   EMBL; AF058889; AAC14346.1; JOINED.
DR   EMBL; AF058890; AAC14346.1; JOINED.
DR   EMBL; AF058891; AAC14346.1; JOINED.
DR   EMBL; AF058892; AAC14346.1; JOINED.
DR   EMBL; AF058893; AAC14346.1; JOINED.
DR   EMBL; AF058894; AAC14346.1; JOINED.
DR   EMBL; AF058895; AAC14346.1; JOINED.
DR   EMBL; AF058896; AAC14346.1; JOINED.
DR   EMBL; AF058897; AAC14346.1; JOINED.
DR   EMBL; AF058898; AAC14346.1; JOINED.
DR   EMBL; AF058899; AAC14346.1; JOINED.
DR   EMBL; AC004257; AAC04897.1; -.
```

```
DR   EMBL; AC004663; AAC15789.1; ALT_INIT.
DR   PIR; S78549; S78549.
DR   HSSP; P00740; 1EDM.
DR   Genew; HGNC:7883; NOTCH3.
DR   MIM; 600276; -.
DR   MIM; 125310; -.
DR   InterPro; IPR002110; ANK.
DR   InterPro; IPR000152; Asx_hydroxyl_S.
DR   InterPro; IPR000742; EGF_2.
DR   InterPro; IPR001881; EGF_Ca.
DR   InterPro; IPR001438; EGF_II.
DR   InterPro; IPR006209; EGF_like.
DR   InterPro; IPR002049; Laminin_EGF.
DR   InterPro; IPR008297; Notch.
DR   InterPro; IPR000800; Notch_dom.
DR   Pfam; PF00023; ank; 6.
DR   Pfam; PF00008; EGF; 34.
DR   Pfam; PF00066; notch; 3.
DR   PIRSF; PIRSF002279; Notch; 1.
DR   PRINTS; PR00010; EGFBLOOD.
DR   PRINTS; PR00011; EGFLAMININ.
DR   PRINTS; PR01452; NOTCH.
DR   SMART; SM00248; ANK; 6.
DR   SMART; SM00179; EGF_CA; 19.
DR   SMART; SM00004; NL; 3.
DR   PROSITE; PS50297; ANK_REP_REGION; 1.
DR   PROSITE; PS50088; ANK_REPEAT; 4.
DR   PROSITE; PS00010; ASX_HYDROXYL; 18.
DR   PROSITE; PS00022; EGF_1; 33.
DR   PROSITE; PS01186; EGF_2; 25.
DR   PROSITE; PS50026; EGF_3; 34.
DR   PROSITE; PS01187; EGF_CA; 16.
KW   Receptor; Transcription regulation; Activator; Differentiation;
KW   Developmental protein; Repeat; ANK repeat; EGF-like domain;
KW   Transmembrane; Glycoprotein; Signal; Phosphorylation; Polymorphism;
KW   Disease mutation.
FT   SIGNAL        1     39       POTENTIAL.
FT   CHAIN        40   2321       NEUROGENIC LOCUS NOTCH HOMOLOG PROTEIN 3.
FT   CHAIN      1629   2321       NOTCH EXTRACELLULAR TRUNCATION (BY
FT                                SIMILARITY).
FT   CHAIN      1662   2321       NOTCH INTRACELLULAR DOMAIN (BY
FT                                SIMILARITY).
FT   DOMAIN       40   1643       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   1644   1664       POTENTIAL.
FT   DOMAIN     1665   2321       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN       40     77       EGF-LIKE 1.
FT   DOMAIN       78    118       EGF-LIKE 2.
FT   DOMAIN      119    156       EGF-LIKE 3.
FT   DOMAIN      158    195       EGF-LIKE 4, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      197    234       EGF-LIKE 5.
FT   DOMAIN      236    272       EGF-LIKE 6, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      274    312       EGF-LIKE 7.
FT   DOMAIN      314    350       EGF-LIKE 8, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      351    389       EGF-LIKE 9.
FT   DOMAIN      391    429       EGF-LIKE 10, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      431    467       EGF-LIKE 11, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      469    505       EGF-LIKE 12, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      507    543       EGF-LIKE 13, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      545    580       EGF-LIKE 14, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      582    618       EGF-LIKE 15, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      620    655       EGF-LIKE 16, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      657    693       EGF-LIKE 17, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      695    730       EGF-LIKE 18.
FT   DOMAIN      734    770       EGF-LIKE 19.
FT   DOMAIN      771    808       EGF-LIKE 20.
FT   DOMAIN      810    847       EGF-LIKE 21, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      849    885       EGF-LIKE 22, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      887    922       EGF-LIKE 23, CALCIUM-BINDING (POTENTIAL).

  Query Match             2.8%;  Score 8;  DB 1;  Length 2321;
  Best Local Similarity   100.0%;  Pred. No. 19;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;
```

```
Qy         62 GFSGPRCQ 69
              ||||||||
Db       1237 GFSGPRCQ 1244


RESULT 13
NTC2_MOUSE
ID   NTC2_MOUSE     STANDARD;      PRT;  2470 AA.
AC   O35516; Q06008; Q60941;
DT   28-FEB-2003 (Rel. 41, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Neurogenic locus notch homolog protein 2 precursor (Notch 2) (Motch
DE   B).
GN   NOTCH2.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6; TISSUE=Thymus;
RA   Hamada Y., Higuchi M., Tsujimoto Y.;
RT   "Complete amino acid sequence and mutliform transcripts encoded by a
RT   single copy of mouse Notch2 gene.";
RL   Submitted (JUL-1994) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE OF 316-1518 FROM N.A.
RC   STRAIN=C57BL/6 X CBA; TISSUE=Embryo;
RX   MEDLINE=93178563; PubMed=8440332;
RA   Lardelli M., Lendahl U.;
RT   "Motch A and Motch B-two mouse Notch homologues coexpressed in a
RT   wide variety of tissues.";
RL   Exp. Cell Res. 204:364-372(1993).
RN   [3]
RP   SEQUENCE OF 1765-2153 FROM N.A.
RX   MEDLINE=97075110; PubMed=8917536;
RA   Milner L.A., Bigas A., Kopan R., Brashem-Stein C., Bernstein I.D.,
RA   Martin D.I.;
RT   "Inhibition of granulocytic differentiation by mNotch1.";
RL   Proc. Natl. Acad. Sci. U.S.A. 93:13014-13019(1996).
RN   [4]
RP   FUNCTION.
RX   MEDLINE=99396706; PubMed=10393120;
RA   Hamada Y., Kadokawa Y., Okabe M., Ikawa M., Coleman J.R.,
RA   Tsujimoto Y.;
RT   "Mutation in ankyrin repeats of the mouse Notch2 gene induces early
RT   embryonic lethality.";
RL   Development 126:3415-3424(1999).
RN   [5]
RP   DEVELOPMENTAL STAGE, AND ALTERNATIVE SPLICING.
RX   MEDLINE=95333893; PubMed=7609614;
RA   Higuchi M., Kiyama H., Hayakawa T., Hamada Y., Tsujimoto Y.;
RT   "Differential expression of Notch1 and Notch2 in developing and adult
RT   mouse brain.";
RL   Brain Res. Mol. Brain Res. 29:263-272(1995).
RN   [6]
RP   POST-TRANSLATIONAL PROCESSING, AND MUTAGENESIS OF MET-1699.
RX   MEDLINE=21523956; PubMed=11518718;
RA   Saxena M.T., Schroeter E.H., Mumm J.S., Kopan R.;
RT   "Murine notch homologs (N1-4) undergo presenilin-dependent
RT   proteolysis.";
RL   J. Biol. Chem. 276:40268-40273(2001).
RN   [7]
RP   POST-TRANSLATIONAL PROCESSING, AND MUTAGENESIS OF MET-1699.
RX   MEDLINE=21374376; PubMed=11459941;
RA   Mizutani T., Taniguchi Y., Aoki T., Hashimoto N., Honjo T.;
RT   "Conservation of the biochemical mechanisms of signal transduction
RT   among mammalian Notch family members.";
RL   Proc. Natl. Acad. Sci. U.S.A. 98:9026-9031(2001).
CC   -!- FUNCTION: Functions as a receptor for membrane-bound ligands
CC       Jagged1, Jagged2 and Delta1 to regulate cell-fate determination.
CC       Upon ligand activation through the released notch intracellular
```

```
CC       domain (NICD) it forms a transcriptional activator complex with
CC       RBP-J kappa and activates genes of the enhancer of split locus.
CC       Affects the implementation of differentiation, proliferation and
CC       apoptotic programs (By similarity). May play an essential role in
CC       postimplantation development, probably in some aspect of cell
CC       specification and/or differentiation.
CC   -!- SUBUNIT: Heterodimer of a C-terminal fragment N(TM) and a N-
CC       terminal fragment N(EC) which are probably linked by disulfide
CC       bonds.
CC   -!- SUBCELLULAR LOCATION: Type I membrane protein. Following
CC       proteolytical processing NICD is translocated to the nucleus.
CC   -!- ALTERNATIVE PRODUCTS:
CC       Event=Alternative splicing; Named isoforms=2;
CC       Name=1;
CC         IsoId=O35516-1; Sequence=Displayed;
CC       Name=2;
CC         IsoId=O35516-2; Sequence=VSP_001405;
CC         Note=No experimental confirmation available;
CC   -!- TISSUE SPECIFICITY: Expressed in the brain, liver, kidney,
CC       neuroepithelia, somites, optic vesicles and branchial arches, but
CC       not heart.
CC   -!- DEVELOPMENTAL STAGE: Expressed in the embryonic ventricular zone,
CC       the postnatal ependymal cells, and the choroid plexus throughout
CC       embryonic and postnatal development.
CC   -!- PTM: Synthesized in the endoplasmic reticulum as an inactive form
CC       which is proteolytically cleaved by a furin-like convertase in the
CC       trans-Golgi network before it reaches the plasma membrane to yield
CC       an active, ligand-accessible form. Cleavage results in a C-
CC       terminal fragment N(TM) and a N-terminal fragment N(EC). Following
CC       ligand binding, it is cleaved by TNF-alpha converting enzyme
CC       (TACE) to yield a membrane-associated intermediate fragment called
CC       notch extracellular truncation (NEXT). This fragment is then
CC       cleaved by presenilin dependent gamma-secretase to release a
CC       notch-derived peptide containing the intracellular domain (NICD)
CC       from the membrane.
CC   -!- PTM: Phosphorylated.
CC   -!- SIMILARITY: Belongs to the NOTCH family.
CC   -!- SIMILARITY: Contains 35 EGF-like domains.
CC   -!- SIMILARITY: Contains 2 Lin/Notch repeats.
CC   -!- SIMILARITY: Contains 6 ANK repeats.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; D32210; BAA22094.1; -.
DR   EMBL; X68279; CAA48340.1; -.
DR   EMBL; U31881; AAC52924.1; -.
DR   PIR; A49175; A49175.
DR   HSSP; P16109; 1FSB.
DR   MGD; MGI:97364; Notch2.
DR   GO; GO:0005887; C:integral to plasma membrane; IC.
DR   GO; GO:0005515; F:protein binding; IPI.
DR   GO; GO:0002011; P:morphogenesis of an epithelial sheet; IMP.
DR   GO; GO:0007219; P:N signaling pathway; IC.
DR   InterPro; IPR002110; ANK.
DR   InterPro; IPR000152; Asx_hydroxyl_S.
DR   InterPro; IPR000742; EGF_2.
DR   InterPro; IPR001881; EGF_Ca.
DR   InterPro; IPR001438; EGF_II.
DR   InterPro; IPR006209; EGF_like.
DR   InterPro; IPR002049; Laminin_EGF.
DR   InterPro; IPR008297; Notch.
DR   InterPro; IPR000800; Notch_dom.
DR   Pfam; PF00023; ank; 6.
DR   Pfam; PF00008; EGF; 34.
DR   Pfam; PF00066; notch; 2.
DR   PIRSF; PIRSF002279; Notch; 1.
DR   PRINTS; PR00010; EGFBLOOD.
DR   PRINTS; PR00011; EGFLAMININ.
DR   PRINTS; PR01452; NOTCH.
DR   SMART; SM00248; ANK; 6.
DR   SMART; SM00179; EGF_CA; 23.
DR   SMART; SM00004; NL; 3.
DR   PROSITE; PS50297; ANK_REP_REGION; 1.
DR   PROSITE; PS50088; ANK_REPEAT; 4.
DR   PROSITE; PS00010; ASX_HYDROXYL; 22.
DR   PROSITE; PS00022; EGF_1; 33.
DR   PROSITE; PS01186; EGF_2; 27.
DR   PROSITE; PS50026; EGF_3; 35.
DR   PROSITE; PS01187; EGF_CA; 22.
KW   Receptor; Transcription regulation; Activator; Differentiation;
KW   Developmental protein; Repeat; ANK repeat; EGF-like domain;
KW   Transmembrane; Glycoprotein; Signal; Phosphorylation;
KW   Alternative splicing.
FT   SIGNAL        1     25       POTENTIAL.
FT   CHAIN        26   2470       NEUROGENIC LOCUS NOTCH HOMOLOG PROTEIN 2.
FT   CHAIN      1666   2470       NOTCH EXTRACELLULAR TRUNCATION.
FT   CHAIN      1697   2470       NOTCH INTRACELLULAR DOMAIN.
FT   DOMAIN       26   1677       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   1678   1698       POTENTIAL.
FT   DOMAIN     1699   2470       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN       26     63       EGF-LIKE 1.
FT   DOMAIN       64    102       EGF-LIKE 2.
FT   DOMAIN      105    143       EGF-LIKE 3.
FT   DOMAIN      144    180       EGF-LIKE 4.
FT   DOMAIN      182    219       EGF-LIKE 5, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      221    256       EGF-LIKE 6 (INCOMPLETE).
FT   DOMAIN      258    294       EGF-LIKE 7, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      296    334       EGF-LIKE 8, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      336    372       EGF-LIKE 9, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      373    411       EGF-LIKE 10.
FT   DOMAIN      413    452       EGF-LIKE 11, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      454    490       EGF-LIKE 12, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      492    528       EGF-LIKE 13, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      530    566       EGF-LIKE 14, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      568    603       EGF-LIKE 15, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      605    641       EGF-LIKE 16, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      643    678       EGF-LIKE 17, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      680    716       EGF-LIKE 18, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      718    753       EGF-LIKE 19.
FT   DOMAIN      755    791       EGF-LIKE 20, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      793    829       EGF-LIKE 21, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      831    869       EGF-LIKE 22.
FT   DOMAIN      871    907       EGF-LIKE 23, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      909    945       EGF-LIKE 24, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      947    983       EGF-LIKE 25, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      985   1021       EGF-LIKE 26, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     1023   1059       EGF-LIKE 27, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     1061   1097       EGF-LIKE 28.
FT   DOMAIN     1099   1145       EGF-LIKE 29.
FT   DOMAIN     1147   1183       EGF-LIKE 30, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     1185   1221       EGF-LIKE 31, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     1223   1260       EGF-LIKE 32, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     1262   1300       EGF-LIKE 33.
FT   DOMAIN     1302   1345       EGF-LIKE 34.
FT   DOMAIN     1372   1410       EGF-LIKE 35.
FT   REPEAT     1418   1454       LIN/NOTCH 1.
FT   REPEAT     1501   1533       LIN/NOTCH 2.
FT   REPEAT     1825   1869       ANK 1.

  Query Match             2.8%;  Score 8;  DB 1;  Length 2470;
  Best Local Similarity   100.0%;  Pred. No. 20;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          5 PCQHGATC 12
              ||||||||
Db       1155 PCQHGATC 1162


RESULT 14
NTC2_HUMAN
```

```
ID   NTC2_HUMAN     STANDARD;      PRT;  2471 AA.
AC   Q04721; Q99734; Q9H240;
DT   28-FEB-2003 (Rel. 41, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Neurogenic locus notch homolog protein 2 precursor (Notch 2) (hN2).
GN   NOTCH2.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Brain;
RA   Blaumueller C.M., Mann R.S.;
RT   "Complete human notch 2 (hN2) cDNA sequence.";
RL   Submitted (NOV-2000) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Breast tumor;
RA   Correa R.G., Camargo A.A., Moreira E.S., Simpson A.J.G.;
RT   "Human Notch2, a novel member of cell-fate determining NOTCH
RT   family.";
RL   Submitted (OCT-2000) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE OF 967-1229 FROM N.A.
RC   TISSUE=T-cell;
RA   Lemasson I., Devaux C., Mesnard J.M.;
RT   "Partial sequence of EGF-like repeat domain of human Notch2 mRNA.";
RL   Submitted (NOV-1996) to the EMBL/GenBank/DDBJ databases.
RN   [4]
RP   SEQUENCE OF 1810-2447 FROM N.A.
RC   TISSUE=Brain;
RX   MEDLINE=93265135; PubMed=1303260;
RA   Stifani S., Blaumueller C.M., Redhead N.J., Hill R.E.,
RA   Artavanis-Tsakonas S.;
RT   "Human homologs of a Drosophila enhancer of split gene product define
RT   a novel family of nuclear proteins.";
RL   Nat. Genet. 2:119-127(1992).
RN   [5]
RP   POST-TRANSLATIONAL PROCESSING.
RX   MEDLINE=97386453; PubMed=9244302;
RA   Blaumueller C.M., Qi H., Zagouras P., Artavanis-Tsakonas S.;
RT   "Intracellular cleavage of Notch leads to a heterodimeric receptor on
RT   the plasma membrane.";
RL   Cell 90:281-291(1997).
RN   [6]
RP   IDENTIFICATION OF LIGANDS.
RX   MEDLINE=99180765; PubMed=10079256;
RA   Gray G.E., Mann R.S., Mitsiadis E., Henrique D., Carcangiu M.-L.,
RA   Banks A., Leiman J., Ward D., Ish-Horowitz D., Artavanis-Tsakonas S.;
RT   "Human ligands of the Notch receptor.";
RL   Am. J. Pathol. 154:785-794(1999).
CC   -!- FUNCTION: Functions as a receptor for membrane-bound ligands
CC       Jagged1, Jagged2 and Delta1 to regulate cell-fate determination.
CC       Upon ligand activation through the released notch intracellular
CC       domain (NICD) it forms a transcriptional activator complex with
CC       RBP-J kappa and activates genes of the enhancer of split locus.
CC       Affects the implementation of differentiation, proliferation and
CC       apoptotic programs (By similarity).
CC   -!- SUBUNIT: Heterodimer of a C-terminal fragment N(TM) and a N-
CC       terminal fragment N(EC) which are probably linked by disulfide
CC       bonds (By similarity).
CC   -!- SUBCELLULAR LOCATION: Type I membrane protein. Following
CC       proteolytical processing NICD is translocated to the nucleus.
CC   -!- TISSUE SPECIFICITY: Expressed in the brain, heart, kidney, lung,
CC       skeletal muscle and liver.
CC   -!- PTM: Synthesized in the endoplasmic reticulum as an inactive form
CC       which is proteolytically cleaved by a furin-like convertase in the
CC       trans-Golgi network before it reaches the plasma membrane to yield
CC       an active, ligand-accessible form. Cleavage results in a C-
CC       terminal fragment N(TM) and a N-terminal fragment N(EC). Following
CC       ligand binding, it is cleaved by TNF-alpha converting enzyme
CC       (TACE) to yield a membrane-associated intermediate fragment called
CC       notch extracellular truncation (NEXT). This fragment is then
CC       cleaved by presenilin dependent gamma-secretase to release a
CC       notch-derived peptide containing the intracellular domain (NICD)
CC       from the membrane (By similarity).
CC   -!- PTM: Phosphorylated (By similarity).
CC   -!- SIMILARITY: Belongs to the NOTCH family.
CC   -!- SIMILARITY: Contains 35 EGF-like domains.
CC   -!- SIMILARITY: Contains 2 Lin/Notch repeats.
CC   -!- SIMILARITY: Contains 6 ANK repeats.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; AF308601; AAA36377.2; -.
DR   EMBL; AF315356; AAG37073.1; -.
DR   EMBL; U77493; AAB19224.1; -.
DR   HSSP; P00740; 1EDM.
DR   Genew; HGNC:7882; NOTCH2.
DR   MIM; 600275; -.
DR   InterPro; IPR002110; ANK.
DR   InterPro; IPR000152; Asx_hydroxyl_S.
DR   InterPro; IPR000742; EGF_2.
DR   InterPro; IPR001881; EGF_Ca.
DR   InterPro; IPR001438; EGF_II.
DR   InterPro; IPR006209; EGF_like.
DR   InterPro; IPR002049; Laminin_EGF.
DR   InterPro; IPR008297; Notch.
DR   InterPro; IPR000800; Notch_dom.
DR   Pfam; PF00023; ank; 6.
DR   Pfam; PF00008; EGF; 35.
DR   Pfam; PF00066; notch; 2.
DR   PIRSF; PIRSF002279; Notch; 1.
DR   PRINTS; PR00010; EGFBLOOD.
DR   PRINTS; PR00011; EGFLAMININ.
DR   PRINTS; PR01452; NOTCH.
DR   SMART; SM00248; ANK; 6.
DR   SMART; SM00179; EGF_CA; 23.
DR   SMART; SM00004; NL; 2.
DR   PROSITE; PS50297; ANK_REP_REGION; 1.
DR   PROSITE; PS50088; ANK_REPEAT; 4.
DR   PROSITE; PS00010; ASX_HYDROXYL; 22.
DR   PROSITE; PS00022; EGF_1; 34.
DR   PROSITE; PS01186; EGF_2; 29.
DR   PROSITE; PS50026; EGF_3; 35.
DR   PROSITE; PS01187; EGF_CA; 22.
KW   Receptor; Transcription regulation; Activator; Differentiation;
KW   Developmental protein; Repeat; ANK repeat; EGF-like domain;
KW   Transmembrane; Glycoprotein; Signal; Phosphorylation.
FT   SIGNAL        1     25       POTENTIAL.
FT   CHAIN        26   2471       NEUROGENIC LOCUS NOTCH HOMOLOG PROTEIN 2.
FT   CHAIN      1666   2471       NOTCH EXTRACELLULAR TRUNCATION (BY
FT                                SIMILARITY).
FT   CHAIN      1697   2471       NOTCH INTRACELLULAR DOMAIN (BY
FT                                SIMILARITY).
FT   DOMAIN       26   1677       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   1678   1698       POTENTIAL.
FT   DOMAIN     1699   2471       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN       26     63       EGF-LIKE 1.
FT   DOMAIN       64    102       EGF-LIKE 2.
FT   DOMAIN      105    143       EGF-LIKE 3.
FT   DOMAIN      144    180       EGF-LIKE 4.
FT   DOMAIN      182    219       EGF-LIKE 5, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      221    258       EGF-LIKE 6.
FT   DOMAIN      260    296       EGF-LIKE 7, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      298    336       EGF-LIKE 8, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      338    374       EGF-LIKE 9, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      375    413       EGF-LIKE 10.
```

```
FT   DOMAIN      415    454       EGF-LIKE 11, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      456    492       EGF-LIKE 12, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      494    530       EGF-LIKE 13, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      532    568       EGF-LIKE 14, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      570    605       EGF-LIKE 15, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      607    643       EGF-LIKE 16, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      645    680       EGF-LIKE 17, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      682    718       EGF-LIKE 18, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      720    755       EGF-LIKE 19.
FT   DOMAIN      757    793       EGF-LIKE 20, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      795    831       EGF-LIKE 21, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      833    871       EGF-LIKE 22.
FT   DOMAIN      873    909       EGF-LIKE 23, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      911    947       EGF-LIKE 24, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      949    985       EGF-LIKE 25, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      987   1023       EGF-LIKE 26, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     1025   1061       EGF-LIKE 27, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     1063   1099       EGF-LIKE 28.
FT   DOMAIN     1101   1147       EGF-LIKE 29.
FT   DOMAIN     1149   1185       EGF-LIKE 30, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     1187   1223       EGF-LIKE 31, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     1225   1262       EGF-LIKE 32, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     1264   1302       EGF-LIKE 33.
FT   DOMAIN     1304   1343       EGF-LIKE 34.
FT   DOMAIN     1374   1412       EGF-LIKE 35.
FT   REPEAT     1420   1456       LIN/NOTCH 1.
FT   REPEAT     1503   1535       LIN/NOTCH 2.
FT   REPEAT     1827   1871       ANK 1.
FT   REPEAT     1876   1905       ANK 2.
FT   REPEAT     1909   1939       ANK 3.
FT   REPEAT     1943   1972       ANK 4.
FT   REPEAT     1976   2005       ANK 5.
FT   REPEAT     2009   2038       ANK 6.
FT   DOMAIN     1645   1648       POLY-ALA.
FT   DOMAIN     1994   1997       POLY-LEU.
FT   DOMAIN     2426   2429       POLY-SER.
FT   DISULFID     28     41       BY SIMILARITY.
FT   DISULFID     35     51       BY SIMILARITY.
FT   DISULFID     53     62       BY SIMILARITY.
FT   DISULFID     68     79       BY SIMILARITY.
FT   DISULFID     73     90       BY SIMILARITY.
FT   DISULFID     92    101       BY SIMILARITY.
FT   DISULFID    109    121       BY SIMILARITY.
FT   DISULFID    115    131       BY SIMILARITY.
FT   DISULFID    133    142       BY SIMILARITY.
FT   DISULFID    148    159       BY SIMILARITY.
FT   DISULFID    153    168       BY SIMILARITY.
FT   DISULFID    170    179       BY SIMILARITY.
FT   DISULFID    186    198       BY SIMILARITY.
FT   DISULFID    192    207       BY SIMILARITY.
FT   DISULFID    209    218       BY SIMILARITY.
FT   DISULFID    225    236       BY SIMILARITY.
FT   DISULFID    230    246       BY SIMILARITY.
FT   DISULFID    248    257       BY SIMILARITY.

  Query Match             2.8%;  Score 8;  DB 1;  Length 2471;
  Best Local Similarity   100.0%;  Pred. No. 20;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          5 PCQHGATC 12
              ||||||||
Db       1157 PCQHGATC 1164


RESULT 15
NTC2_RAT
ID   NTC2_RAT       STANDARD;      PRT;  2471 AA.
AC   Q9QW30;
DT   28-FEB-2003 (Rel. 41, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Neurogenic locus notch homolog protein 2 precursor (Notch 2).
GN   NOTCH2.
```

```
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Brain;
RX   MEDLINE=93202015; PubMed=1295745;
RA   Weinmaster G., Roberts V.J., Lemke G.;
RT   "Notch2: a second mammalian Notch gene.";
RL   Development 116:931-941(1992).
RN   [2]
RP   TISSUE SPECIFICITY.
RX   MEDLINE=21331789; PubMed=11438922;
RA   Irvin D.K., Zurcher S.D., Nguyen T., Weinmaster G., Kornblum H.I.;
RT   "Expression patterns of Notch1, Notch2, and Notch3 suggest multiple
RT   functional roles for the Notch-DSL signaling system during brain
RT   development.";
RL   J. Comp. Neurol. 436:167-181(2001).
CC   -!- FUNCTION: Functions as a receptor for membrane-bound ligands
CC       Jagged1, Jagged2 and Delta1 to regulate cell-fate determination.
CC       Upon ligand activation through the released notch intracellular
CC       domain (NICD) it forms a transcriptional activator complex with
CC       RBP-J kappa and activates genes of the enhancer of split locus.
CC       Affects the implementation of differentiation, proliferation and
CC       apoptotic programs. May play an essential role in postimplantation
CC       development, probably in some aspect of cell specification and/or
CC       differentiation (By similarity).
CC   -!- SUBUNIT: Heterodimer of a C-terminal fragment N(TM) and a N-
CC       terminal fragment N(EC) which are probably linked by disulfide
CC       bonds (By similarity).
CC   -!- SUBCELLULAR LOCATION: Type I membrane protein. Following
CC       proteolytical processing NICD is translocated to the nucleus.
CC   -!- TISSUE SPECIFICITY: Highly expressed in the spleen and choroid
CC       plexus in the brain. Expressed in postnatal central nervous system
CC       (CNS) germinal zones and, in early postnatal life, within numerous
CC       cells throughout the CNS. It is more highly localized to
CC       ventricular germinal zones. Also found in the heart, liver and
CC       kidney.
CC   -!- DEVELOPMENTAL STAGE: Expressed in the brain during E14 and E17.
CC   -!- PTM: Synthesized in the endoplasmic reticulum as an inactive form
CC       which is proteolytically cleaved by a furin-like convertase in the
CC       trans-Golgi network before it reaches the plasma membrane to yield
CC       an active, ligand-accessible form. Cleavage results in a C-
CC       terminal fragment N(TM) and a N-terminal fragment N(EC). Following
CC       ligand binding, it is cleaved by TNF-alpha converting enzyme
CC       (TACE) to yield a membrane-associated intermediate fragment called
CC       notch extracellular truncation (NEXT). This fragment is then
CC       cleaved by presenilin dependent gamma-secretase to release a
CC       notch-derived peptide containing the intracellular domain (NICD)
CC       from the membrane (By similarity).
CC   -!- PTM: Phosphorylated (By similarity).
CC   -!- SIMILARITY: Belongs to the NOTCH family.
CC   -!- SIMILARITY: Contains 35 EGF-like domains.
CC   -!- SIMILARITY: Contains 2 Lin/Notch repeats.
CC   -!- SIMILARITY: Contains 6 ANK repeats.
CC
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; M93661; AAK13558.1; -.
DR   PIR; A49128; A49128.
DR   HSSP; P00743; 1CCF.
DR   InterPro; IPR002110; ANK.
DR   InterPro; IPR000152; Asx_hydroxyl_S.
DR   InterPro; IPR000742; EGF_2.
DR   InterPro; IPR001881; EGF_Ca.
DR   InterPro; IPR001438; EGF_II.
```

```
DR   InterPro; IPR006209; EGF_like.
DR   InterPro; IPR002049; Laminin_EGF.
DR   InterPro; IPR008297; Notch.
DR   InterPro; IPR000800; Notch_dom.
DR   Pfam; PF00023; ank; 6.
DR   Pfam; PF00008; EGF; 35.
DR   Pfam; PF00066; notch; 2.
DR   PIRSF; PIRSF002279; Notch; 1.
DR   PRINTS; PR00010; EGFBLOOD.
DR   PRINTS; PR00011; EGFLAMININ.
DR   PRINTS; PR01452; NOTCH.
DR   SMART; SM00248; ANK; 6.
DR   SMART; SM00179; EGF_CA; 24.
DR   SMART; SM00004; NL; 2.
DR   PROSITE; PS50297; ANK_REP_REGION; 1.
DR   PROSITE; PS50088; ANK_REPEAT; 4.
DR   PROSITE; PS00010; ASX_HYDROXYL; 22.
DR   PROSITE; PS00022; EGF_1; 34.
DR   PROSITE; PS01186; EGF_2; 26.
DR   PROSITE; PS50026; EGF_3; 35.
DR   PROSITE; PS01187; EGF_CA; 22.
KW   Receptor; Transcription regulation; Activator; Differentiation;
KW   Developmental protein; Repeat; ANK repeat; EGF-like domain;
KW   Transmembrane; Glycoprotein; Signal; Phosphorylation.
FT   SIGNAL        1     25       POTENTIAL.
FT   CHAIN        26   2471       NEUROGENIC LOCUS NOTCH HOMOLOG PROTEIN 2.
FT   CHAIN      1666   2471       NOTCH EXTRACELLULAR TRUNCATION (BY
FT                                SIMILARITY).
FT   CHAIN      1697   2471       NOTCH INTRACELLULAR DOMAIN (BY
FT                                SIMILARITY).
FT   DOMAIN       26   1677       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   1678   1698       POTENTIAL.
FT   DOMAIN     1699   2471       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN       26     63       EGF-LIKE 1.
FT   DOMAIN       64    102       EGF-LIKE 2.
FT   DOMAIN      105    143       EGF-LIKE 3.
FT   DOMAIN      144    180       EGF-LIKE 4.
FT   DOMAIN      182    219       EGF-LIKE 5, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      221    258       EGF-LIKE 6.
FT   DOMAIN      260    296       EGF-LIKE 7, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      298    336       EGF-LIKE 8, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      338    374       EGF-LIKE 9, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      375    413       EGF-LIKE 10.
FT   DOMAIN      415    454       EGF-LIKE 11, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      456    492       EGF-LIKE 12, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      494    530       EGF-LIKE 13, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      532    568       EGF-LIKE 14, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      570    605       EGF-LIKE 15, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      607    643       EGF-LIKE 16, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      645    680       EGF-LIKE 17, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      682    718       EGF-LIKE 18, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      720    755       EGF-LIKE 19.
FT   DOMAIN      757    793       EGF-LIKE 20, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      795    831       EGF-LIKE 21, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      833    871       EGF-LIKE 22.
FT   DOMAIN      873    909       EGF-LIKE 23, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      911    947       EGF-LIKE 24, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      949    985       EGF-LIKE 25, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      987   1023       EGF-LIKE 26, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     1025   1061       EGF-LIKE 27, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     1063   1099       EGF-LIKE 28.
FT   DOMAIN     1101   1147       EGF-LIKE 29.
FT   DOMAIN     1149   1185       EGF-LIKE 30, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     1187   1223       EGF-LIKE 31, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     1225   1262       EGF-LIKE 32, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     1264   1302       EGF-LIKE 33.
FT   DOMAIN     1304   1343       EGF-LIKE 34.
FT   DOMAIN     1374   1412       EGF-LIKE 35.
FT   DOMAIN     1645   1648       POLY-ALA.
FT   DOMAIN     1994   1997       POLY-LEU.
FT   DOMAIN     2426   2429       POLY-SER.
FT   DOMAIN     2446   2451       POLY-GLY.
FT   REPEAT     1420   1456       LIN/NOTCH 1.
FT   REPEAT     1503   1535       LIN/NOTCH 2.
FT   REPEAT     1827   1871       ANK 1.
FT   REPEAT     1876   1905       ANK 2.
FT   REPEAT     1909   1939       ANK 3.
FT   REPEAT     1943   1972       ANK 4.
FT   REPEAT     1976   2005       ANK 5.
FT   REPEAT     2009   2038       ANK 6.
FT   DISULFID     28     41       BY SIMILARITY.
FT   DISULFID     35     51       BY SIMILARITY.
FT   DISULFID     53     62       BY SIMILARITY.
FT   DISULFID     68     79       BY SIMILARITY.
FT   DISULFID     73     90       BY SIMILARITY.
FT   DISULFID     92    101       BY SIMILARITY.
FT   DISULFID    109    121       BY SIMILARITY.
FT   DISULFID    115    131       BY SIMILARITY.
FT   DISULFID    133    142       BY SIMILARITY.
FT   DISULFID    148    159       BY SIMILARITY.
FT   DISULFID    153    168       BY SIMILARITY.
FT   DISULFID    170    179       BY SIMILARITY.
FT   DISULFID    186    198       BY SIMILARITY.
FT   DISULFID    192    207       BY SIMILARITY.
FT   DISULFID    209    218       BY SIMILARITY.
FT   DISULFID    225    236       BY SIMILARITY.
FT   DISULFID    230    246       BY SIMILARITY.
FT   DISULFID    248    257       BY SIMILARITY.
FT   DISULFID    264    275       BY SIMILARITY.
FT   DISULFID    269    284       BY SIMILARITY.
FT   DISULFID    286    295       BY SIMILARITY.
FT   DISULFID    302    315       BY SIMILARITY.
FT   DISULFID    309    324       BY SIMILARITY.
FT   DISULFID    326    335       BY SIMILARITY.
FT   DISULFID    342    353       BY SIMILARITY.
FT   DISULFID    347    362       BY SIMILARITY.
FT   DISULFID    364    373       BY SIMILARITY.
FT   DISULFID    379    390       BY SIMILARITY.
FT   DISULFID    384    401       BY SIMILARITY.
FT   DISULFID    403    412       BY SIMILARITY.
FT   DISULFID    419    433       BY SIMILARITY.
FT   DISULFID    427    442       BY SIMILARITY.
FT   DISULFID    444    453       BY SIMILARITY.
FT   DISULFID    460    471       BY SIMILARITY.
FT   DISULFID    465    480       BY SIMILARITY.
FT   DISULFID    482    491       BY SIMILARITY.
FT   DISULFID    498    509       BY SIMILARITY.
FT   DISULFID    503    518       BY SIMILARITY.
FT   DISULFID    520    529       BY SIMILARITY.
FT   DISULFID    536    547       BY SIMILARITY.

  Query Match             2.8%;  Score 8;  DB 1;  Length 2471;
  Best Local Similarity   100.0%;  Pred. No. 20;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          5 PCQHGATC 12
              ||||||||
Db       1157 PCQHGATC 1164

Search completed: March  9, 2004, 17:28:35
Job time : 7.86906 secs
```

```
                    GenCore version 5.1.6
           Copyright (c) 1993 - 2004  Compugen Ltd.

OM protein - protein search, using sw model

Run on:         March  9, 2004, 17:22:43 ; Search time 24.8707 Seconds
                                           (without alignments)
                                           3602.917 Million cell updates/sec

Title:          US-10-006-011A-9
Perfect score:  284
Sequence:       1 CERQPCQHGATCMPAGEYEF..........QPLDLQHRAQAGANTRPCPS 284

Scoring table:  OLIGO
                Gapop 60.0 , Gapext 60.0

Searched:       1017041 seqs, 315518202 residues

Word size :      0

Total number of hits satisfying chosen parameters:      1017041

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :       SPTREMBL_25:*
                1:  sp_archea:*
                2:  sp_bacteria:*
                3:  sp_fungi:*
                4:  sp_human:*
                5:  sp_invertebrate:*
                6:  sp_mammal:*
                7:  sp_mhc:*
                8:  sp_organelle:*
                9:  sp_phage:*
                10:  sp_plant:*
                11:  sp_rodent:*
                12:  sp_virus:*
                13:  sp_vertebrate:*
                14:  sp_unclassified:*
                15:  sp_rvirus:*
                16:  sp_bacteriap:*
                17:  sp_archeap:*

      Pred. No. is the number of results predicted by chance to have a
      score greater than or equal to the score of the result being printed,
      and is derived by analysis of the total score distribution.

                                SUMMARIES
                     %
Result             Query
  No.    Score   Match Length DB  ID                        Description
  ------------------------------------------------------------------------
    1       13     4.6     68  11  O08591                    O08591 rattus norv
    2       10     3.5     53  13  Q9DFL0                    Q9df10 gillichthys
    3        9     3.2   1208  11  Q80YA8                    Q80ya8 mus musculu
    4        9     3.2   2516  11  Q7TQ52                    Q7tq52 mus musculu
    5        9     3.2   2526  11  Q7TQ51                    Q7tq51 mus musculu
    6        9     3.2   2531  11  Q8K428                    Q8k428 mus musculu
    7        9     3.2   2531  11  Q7TQ50                    Q7tq50 mus musculu
    8        9     3.2   3367  5  Q9XZC9                    Q9xzc9 drosophila
    9        9     3.2   3375  5  Q8IP51                    Q8ip51 drosophila
   10        8     2.8    219  11  Q8BMI5                    Q8bmi5 mus musculu
   11        8     2.8    494  16  Q81XS5                    Q81xs5 bacillus an
   12        8     2.8    494  16  Q816E3                    Q816e3 bacillus ce
   13        8     2.8    497  2  Q7X2C8                    Q7x2c8 geobacillus
   14        8     2.8    497  16  Q8YWD5                    Q8ywd5 anabaena sp
   15        8     2.8    615  13  O57409                    O57409 brachydanio
   16        8     2.8    642  13  P79941                    P79941 xenopus lae
   17        8     2.8    642  13  Q7ZXT4                    Q7zxt4 xenopus lae
   18        8     2.8    659  11  Q7TQI9                    Q7tqi9 mus musculu
   19        8     2.8    816  11  O70474                    O70474 rattus norv
   20        8     2.8    858  11  Q8R145                    Q8r145 mus musculu
   21        8     2.8   1290  13  Q9W6E1                    Q9w6e1 gallus gall
   22        8     2.8   1551  5  Q9NGV4                    Q9ngv4 drosophila
   23        8     2.8   2447  13  O13149                    O13149 fugu rubrip
   24        8     2.8   3396  5  Q9VM55                    Q9vm55 drosophila
   25        7     2.5     49  8  O79191                    O79191 ninox stren
   26        7     2.5     71  16  Q7UXR0                    Q7uxr0 rhodopirell
   27        7     2.5     91  16  Q82PU0                    Q82pu0 streptomyce
   28        7     2.5     93  16  P74056                    P74056 synechocyst
   29        7     2.5    106  4  Q9UN93                    Q9un93 homo sapien
   30        7     2.5    106  10  Q9XHS4                    Q9xhs4 brassica ca
   31        7     2.5    108  10  Q9XHS6                    Q9xhs6 brassica ol
   32        7     2.5    122  4  Q8N8V6                    Q8n8v6 homo sapien
   33        7     2.5    151  12  Q918M4                    Q918m4 measles vir
   34        7     2.5    152  4  Q8N124                    Q8n124 homo sapien
   35        7     2.5    152  4  Q8N197                    Q8n197 homo sapien
   36        7     2.5    153  5  Q9GPR3                    Q9gpr3 dictyosteli
   37        7     2.5    157  16  Q82F13                    Q82f13 streptomyce
   38        7     2.5    162  16  Q7UDF5                    Q7udf5 shigella fl
   39        7     2.5    165  16  Q7WM03                    Q7wm03 bordetella
   40        7     2.5    172  16  Q7W8E2                    Q7w8e2 bordetella
   41        7     2.5    181  16  Q9A8I9                    Q9a8i9 caulobacter
   42        7     2.5    185  2  Q8KP10                    Q8kp10 bacillus me
   43        7     2.5    185  11  Q8BKJ6                    Q8bkj6 mus musculu
   44        7     2.5    186  5  Q8MSM1                    Q8msm1 drosophila
   45        7     2.5    193  11  Q8VD50                    Q8vd50 rattus norv

                              ALIGNMENTS

RESULT 1
O08591
ID   O08591      PRELIMINARY;      PRT;    68 AA.
AC   O08591;
DT   01-JUL-1997 (TrEMBLrel. 04, Created)
DT   01-JUL-1997 (TrEMBLrel. 04, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Perlecan (Fragment).
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Kidney;
RA   Pyke C., Kristensen P., Ostergaard P.B., Oturai P.S., Romer J.;
RT   "Proteoglycan expression in the normal rat kidney.";
RL   Submitted (OCT-1996) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; U75305; AAB51124.1; -.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR001791; Laminin_G.
DR   Pfam; PF00054; laminin_G; 1.
DR   PROSITE; PS50025; LAM_G_DOMAIN; 1.
FT   NON_TER       1       1
FT   NON_TER      68      68
SQ   SEQUENCE   68 AA;  7395 MW;  5868E45D8A7083E0 CRC64;

  Query Match             4.6%;  Score 13;  DB 11;  Length 68;
  Best Local Similarity   100.0%;  Pred. No. 1.9e-05;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        112 FSRSLPEVPETIE 124
              |||||||||||||
Db         21 FSRSLPEVPETIE 33


RESULT 2
Q9DFL0
ID   Q9DFL0      PRELIMINARY;      PRT;    53 AA.
```

```
AC   Q9DFL0;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   S-adenosylmethionine synthase-like protein (Fragment).
OS   Gillichthys mirabilis (Long-jawed mudsucker).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC   Acanthomorpha; Acanthopterygii; Percomorpha; Perciformes; Gobioidei;
OC   Gobiidae; Gillichthys.
OX   NCBI_TaxID=8222;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Liver;
RX   MEDLINE=21117151; PubMed=11172064;
RA   Gracey A.Y., Troll J.V., Somero G.N.;
RT   "Hypoxia-induced gene expression profiling in the euryoxic fish
RT   Gillichthys mirabilis.";
RL   Proc. Natl. Acad. Sci. U.S.A. 98:1993-1998(2001).
DR   EMBL; AF266225; AAG13345.1; -.
FT   NON_TER      1      1
SQ   SEQUENCE   53 AA;  5541 MW;  A7D6F0FF411CF46E CRC64;

  Query Match            3.5%;  Score 10;  DB 13;  Length 53;
  Best Local Similarity  100.0%;  Pred. No. 0.023;
  Matches   10;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         53 QGTRCLCLPG 62
              ||||||||||
Db         35 QGTRCLCLPG 44

RESULT 3
Q80YA8
ID   Q80YA8      PRELIMINARY;      PRT;  1208 AA.
AC   Q80YA8;
DT   01-JUN-2003 (TrEMBLrel. 24, Created)
DT   01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   5930402A21 protein (Fragment).
GN   5930402A21.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6; TISSUE=Brain;
RX   MEDLINE=22388257; PubMed=12477932;
RA   Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA   Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA   Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA   Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA   Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA   Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA   Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA   Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA   Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA   Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA   Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA   Fahey J., Helton E., Ketteman M., Madan A., Rodrigues S., Sanchez A.,
RA   Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA   Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA   Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA   Krzywinski M.I., Skalska U., Smailus D.E., Schnerch A., Schein J.E.,
RA   Jones S.J., Marra M.A.;
RT   "Generation and initial analysis of more than 15,000 full-length human
RT   and mouse cDNA sequences.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6; TISSUE=Brain;
RA   Strausberg R.;
RL   Submitted (JAN-2003) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; BC043114; AAH43114.1; -.
DR   GO; GO:0005509; F:calcium ion binding; IEA.
DR   GO; GO:0005198; F:structural molecule activity; IEA.
DR   InterPro; IPR000152; Asx_hydroxyl_S.
DR   InterPro; IPR000742; EGF_2.
DR   InterPro; IPR001881; EGF_Ca.
DR   InterPro; IPR001438; EGF_II.
DR   InterPro; IPR006209; EGF_like.
DR   InterPro; IPR006210; IEGF.
DR   InterPro; IPR002049; Laminin_EGF.
DR   InterPro; IPR001791; Laminin_G.
DR   Pfam; PF00008; EGF; 13.
DR   PRINTS; PR00010; EGFBLOOD.
DR   PRINTS; PR00011; EGFLAMININ.
DR   SMART; SM00181; EGF; 15.
DR   SMART; SM00179; EGF_CA; 13.
DR   SMART; SM00282; LamG; 3.
DR   PROSITE; PS00010; ASX_HYDROXYL; 8.
DR   PROSITE; PS00022; EGF_1; 14.
DR   PROSITE; PS01186; EGF_2; 9.
DR   PROSITE; PS01187; EGF_CA; 5.
DR   PROSITE; PS50025; LAM_G_DOMAIN; 2.
FT   NON_TER      1      1
SQ   SEQUENCE   1208 AA;  126978 MW;  FDFF2DF2F0B2F198 CRC64;

  Query Match            3.2%;  Score 9;  DB 11;  Length 1208;
  Best Local Similarity  100.0%;  Pred. No. 4.3;
  Matches    9;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         44 EPCLHGGTC 52
              |||||||||
Db        295 EPCLHGGTC 303

RESULT 4
Q7TQ52
ID   Q7TQ52      PRELIMINARY;      PRT;  2516 AA.
AC   Q7TQ52;
DT   01-OCT-2003 (TrEMBLrel. 25, Created)
DT   01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Transmembrane receptor Notch1 B.
GN   NOTCH1.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C.B-17; TISSUE=Thymus;
RA   Tsuji H., Ishii-Ohba H., Ukai H., Katsube T., Ogiu T.;
RT   "Radiation-induced deletions in the 5' end region of Notch1 lead to
RT   the formation of truncated proteins and are involved in the
RT   development of mouse thymic lymphomas.";
RL   Carcinogenesis 24:1-12(2003).
DR   EMBL; AB100603; BAC77038.1; -.
KW   Receptor; Transmembrane.
SQ   SEQUENCE   2516 AA;  269177 MW;  17FD72740EBD6E35 CRC64;

  Query Match            3.2%;  Score 9;  DB 11;  Length 2516;
  Best Local Similarity  100.0%;  Pred. No. 8.3;
  Matches    9;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         45 PCLHGGTCQ 53
              |||||||||
Db       1014 PCLHGGTCQ 1022

RESULT 5
Q7TQ51
ID   Q7TQ51      PRELIMINARY;      PRT;  2526 AA.
AC   Q7TQ51;
```

```
DT   01-OCT-2003 (TrEMBLrel. 25, Created)
DT   01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Transmembrane receptor Notch1 D.
GN   NOTCH1.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C.B-17; TISSUE=Thymus;
RA   Tsuji H., Ishii-Ohba H., Ukai H., Katsube T., Ogiu T.;
RT   "Radiation-induced deletions in the 5' end region of Notch1 lead to
RT   the formation of truncated proteins and are involved in the
RT   development of mouse thymic lymphomas.";
RL   Carcinogenesis 24:1-12(2003).
DR   EMBL; AB100603; BAC77039.1; -.
KW   Receptor; Transmembrane.
SQ   SEQUENCE   2526 AA;  270583 MW;  017563FCE9703264 CRC64;

  Query Match             3.2%;  Score 9;  DB 11;  Length 2526;
  Best Local Similarity   100.0%;  Pred. No. 8.4;
  Matches    9;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         45 PCLHGGTCQ 53
              |||||||||
Db       1024 PCLHGGTCQ 1032


RESULT 6
Q8K428
ID   Q8K428      PRELIMINARY;      PRT;  2531 AA.
AC   Q8K428;
DT   01-OCT-2002 (TrEMBLrel. 22, Created)
DT   01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Notch 1 protein.
GN   NOTCH1.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BALB/c; TISSUE=Thymus;
RX   MEDLINE=95044925; PubMed=7956822;
RA   Nye J.S., Kopan R., Axel R.;
RT   "An activated Notch suppresses neurogenesis and myogenesis but not
RT   gliogenesis in mammalian cells.";
RL   Development 120:2421-2430(1994).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BALB/c; TISSUE=Thymus;
RX   MEDLINE=22119593; PubMed=12123574;
RA   Foltz D.R., Santiago M.C., Berechid B.E., Nye J.S.;
RT   "Glycogen Synthase Kinase-3beta Modulates Notch Signaling and
RT   Stability.";
RL   Curr. Biol. 12:1006-1011(2002).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BALB/c; TISSUE=Thymus;
RA   Schellin K.A., Pauley A.M., Nye J.S.;
RL   Submitted (MAY-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF508809; AAM28905.1; -.
DR   MGD; MGI:97363; Notch1.
DR   GO; GO:0005515; F:protein binding; IPI.
DR   GO; GO:0030154; P:cell differentiation; IMP.
DR   GO; GO:0007386; P:compartment specification; IMP.
DR   GO; GO:0045944; P:positive regulation of transcription from P. . .; IDA.
DR   InterPro; IPR002110; ANK.
DR   InterPro; IPR000152; Asx_hydroxyl_S.
DR   InterPro; IPR000742; EGF_2.
```

```
DR   InterPro; IPR001881; EGF_Ca.
DR   InterPro; IPR001438; EGF_II.
DR   InterPro; IPR006209; EGF_like.
DR   InterPro; IPR006210; IEGF.
DR   InterPro; IPR002049; Laminin_EGF.
DR   InterPro; IPR008297; Notch.
DR   InterPro; IPR000800; Notch_dom.
DR   Pfam; PF00023; ank; 6.
DR   Pfam; PF00008; EGF; 36.
DR   Pfam; PF00066; notch; 3.
DR   PRINTS; PR00010; EGFBLOOD.
DR   PRINTS; PR00011; EGFLAMININ.
DR   PRINTS; PR01452; NOTCH.
DR   SMART; SM00248; ANK; 6.
DR   SMART; SM00181; EGF; 37.
DR   SMART; SM00179; EGF_CA; 35.
DR   SMART; SM00004; NL; 3.
DR   PROSITE; PS50088; ANK_REPEAT; 4.
DR   PROSITE; PS50297; ANK_REP_REGION; 1.
DR   PROSITE; PS00010; ASX_HYDROXYL; 22.
DR   PROSITE; PS00022; EGF_1; 35.
DR   PROSITE; PS01186; EGF_2; 27.
DR   PROSITE; PS01187; EGF_CA; 21.
DR   PIRSF; PIRSF002279; Notch; 1.
KW   ANK repeat; EGF-like domain; Repeat.
SQ   SEQUENCE   2531 AA;  270819 MW;  7DB7E0DEF799D999 CRC64;

  Query Match             3.2%;  Score 9;  DB 11;  Length 2531;
  Best Local Similarity   100.0%;  Pred. No. 8.4;
  Matches    9;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         45 PCLHGGTCQ 53
              |||||||||
Db       1029 PCLHGGTCQ 1037


RESULT 7
Q7TQ50
ID   Q7TQ50      PRELIMINARY;      PRT;  2531 AA.
AC   Q7TQ50;
DT   01-OCT-2003 (TrEMBLrel. 25, Created)
DT   01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Transmembrane receptor Notch1.
GN   NOTCH1.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C.B-17; TISSUE=Thymus;
RA   Tsuji H., Ishii-Ohba H., Ukai H., Katsube T., Ogiu T.;
RT   "Radiation-induced deletions in the 5' end region of Notch1 lead to
RT   the formation of truncated proteins and are involved in the
RT   development of mouse thymic lymphomas.";
RL   Carcinogenesis 24:1-12(2003).
DR   EMBL; AB100603; BAC77040.1; -.
KW   Receptor; Transmembrane.
SQ   SEQUENCE   2531 AA;  270832 MW;  97C91F69BABF02BF CRC64;

  Query Match             3.2%;  Score 9;  DB 11;  Length 2531;
  Best Local Similarity   100.0%;  Pred. No. 8.4;
  Matches    9;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         45 PCLHGGTCQ 53
              |||||||||
Db       1029 PCLHGGTCQ 1037


RESULT 8
Q9XZC9
ID   Q9XZC9      PRELIMINARY;      PRT;  3367 AA.
```

```
AC   Q9XZC9;
DT   01-NOV-1999 (TrEMBLrel. 12, Created)
DT   01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Laminin ALPHA1,2 (SYMBOL=WB).
GN   WB OR WING BLISTER OR CG15288.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC   Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=OREGON-R;
RX   MEDLINE=99207061; PubMed=10189378;
RA   Martin D., Zusman S., Li X., Williams E.L., Khare N., DaRocha S.,
RA   Chiquet-Ehrismann R., Baumgartner S.;
RT   "wing blister, a new Drosophila laminin alpha chain required for cell
RT   adhesion and migration during embryonic and imaginal development.";
RL   J. Cell Biol. 145:191-201(1999).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Berkeley;
RX   MEDLINE=99403001; PubMed=10471707;
RA   Ashburner M., Misra S., Roote J., Lewis S.E., Blazej R., Davis T.,
RA   Doyle C., Galle R., George R., Harris N., Hartzell G., Harvey D.,
RA   Hong L., Houston K., Hoskins R., Johnson G., Martin C., Moshrefi A.,
RA   Palazzolo M., Reese M.G., Spradling A., Tsang G., Wan K., Whitelaw K.,
RA   Celniker S., Rubin G.M.;
RT   "An exploration of the sequence of a 2.9-Mb region of the genome of
RT   Drosophila melanogaster: the Adh region.";
RL   Genetics 153:179-219(1999).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Berkeley;
RA   Celniker S.E., Agbayani A., Arcaina T.T., Baxter E., Blazej R.G.,
RA   Butenhoff C., Champe M., Chavez C., Chew M., Ciesiolka L., Doyle C.M.,
RA   Farfan D.E., Galle R., George R.A., Harris N.L., Hoskins R.A.,
RA   Houston K.A., Hummasti S.R., Karra K., Kearney L., Kim E., Lee B.,
RA   Lewis S., Li P., Lomotan M.A., Mazda P., Moshrefi A.R., Moshrefi M.,
RA   Nixon K., Pacleb J.M., Park S., Pfeiffer B., Poon L., Sequeira A.,
RA   Sethi H., Snir E., Svirskas R.R., Wan K.H., Weinburg T., Zhang R.,
RA   Zieran L.L., Rubin G.M.;
RL   Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF135118; AAD31714.1; -.
DR   EMBL; AE003409; AAF44861.1; -.
DR   HSSP; P00740; 1EDM.
DR   FlyBase; FBgn0004002; wb.
DR   GO; GO:0007267; P:cell-cell signaling; NAS.
DR   InterPro; IPR001589; Actbind_actnin.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR006209; EGF_like.
DR   InterPro; IPR000034; Laminin_B.
DR   InterPro; IPR002049; Laminin_EGF.
DR   InterPro; IPR001791; Laminin_G.
DR   InterPro; IPR008211; LamNT.
DR   Pfam; PF00052; laminin_B; 2.
DR   Pfam; PF00053; laminin_EGF; 14.
DR   Pfam; PF00054; laminin_G; 4.
DR   Pfam; PF00055; laminin_Nterm; 1.
DR   PRINTS; PR00011; EGFLAMININ.
DR   SMART; SM00180; EGF_Lam; 16.
DR   SMART; SM00282; LamG; 5.
DR   SMART; SM00136; LamNT; 1.
DR   PROSITE; PS00019; ACTININ_1; 1.
DR   PROSITE; PS00022; EGF_1; 14.
DR   PROSITE; PS01186; EGF_2; 4.
DR   PROSITE; PS01248; LAMININ_TYPE_EGF; 16.
DR   PROSITE; PS50025; LAM_G_DOMAIN; 4.
KW   Laminin EGF-like domain.
SQ   SEQUENCE   3367 AA;  374097 MW;  EB125654B1BC1511 CRC64;

  Query Match             3.2%;  Score 9;  DB 5;  Length 3367;
  Best Local Similarity   100.0%;  Pred. No. 11;
  Matches    9;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        182 NDGEWHRVT 190
              |||||||||
Db       3081 NDGEWHRVT 3089


RESULT 9
Q8IP51
ID   Q8IP51      PRELIMINARY;      PRT;  3375 AA.
AC   Q8IP51;
DT   01-MAR-2003 (TrEMBLrel. 23, Created)
DT   01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   CG15288-PB.
GN   WB OR CG15288.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC   Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=20196006; PubMed=10731132;
RA   Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA   Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA   George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA   Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA   Brandon R.C., Rogers Y.H., Blazej R.G., Champe M., Pfeiffer B.D.,
RA   Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Gabor G.L.,
RA   Abril J.F., Agbayani A., An H.J., Andrews-Pfannkoch C., Baldwin D.,
RA   Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA   Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA   Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA   Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA   Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA   de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA   Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA   Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA   Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA   Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA   Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA   Hostin D., Houston K.A., Howland T.J., Wei M.H., Ibegwam C.,
RA   Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA   Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA   Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA   Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA   Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA   Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA   Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA   Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA   Reinert K., Remington K., Saunders R.D., Scheeler F., Shen H.,
RA   Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA   Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA   Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA   Wang Z.Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA   Williams S.M., WoodageT, Worley K.C., Wu D., Yang S., Yao Q.A., Ye J.,
RA   Yeh R.F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA   Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA   Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT   "The genome sequence of Drosophila melanogaster.";
RL   Science 287:2185-2195(2000).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Celniker S.E., Adams M.D., Kronmiller B., Wan K.H., Holt R.A.,
RA   Evans C.A., Gocayne J.D., Amanatides P.G., Brandon R.C., Rogers Y.,
RA   Banzon J., An H., Baldwin D., Banzon J., Beeson K.Y., Busam D.A.,
RA   Carlson J.W., Center A., Champe M., Davenport L.B., Dietz S.M.,
RA   Dodson K., Dorsett V., Doup L.E., Doyle C., Dresnek D., Farfan D.,
RA   Ferriera S., Frise E., Galle R.F., Garg N.S., George R.A.,
RA   Gonzalez M., Houck J., Hoskins R.A., Hostin D., Howland T.J.,
RA   Ibegwam C., Jalali M., Kruse D., Li P., Mattei B., Moshrefi A.,
```

```
RA   McIntosh T.C., Moy M., Murphy B., Nelson C., Nelson K.A., Nunoo J.,
RA   Pacleb J., Paragas V., Park S., Patel S., Pfeiffer B.,
RA   Phouanenavong S., Pittman G.S., Puri V., Richards S., Scheeler F.,
RA   Stapleton M., Strong R., Svirskas R., Tector C., Tyler D.,
RA   Williams S.M., Zaveri J.S., Smith H.O., Venter J.C., Rubin G.M.;
RT   "Sequencing of Drosophila melanogaster genome.";
RL   Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RA   Misra S., Crosby M.A., Matthews B.B., Bayraktaroglu L., Campbell K.,
RA   Hradecky P., Huang Y., Kaminker J.S., Prochnik S.E., Smith C.D.,
RA   Tupy J.L., Bergman C., Berman B., Carlson J.W., Celniker S.E.,
RA   Clamp M., Drysdale R., Emmert D., Frise E., de Grey A., Harris N.,
RA   Kronmiller B., Marshall B., Millburn G., Richter J., Russo S.,
RA   Searle S.M.J., Smith E., Shu S., Smutniak F., Whitfield E.,
RA   Ashburner M., Gelbart W.M., Rubin G.M., Mungall C.J., Lewis S.E.;
RT   "Annotation of Drosophila melanogaster genome.";
RL   Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
RN   [4]
RP   SEQUENCE FROM N.A.
RA   Adams M.D., Celniker S.E., Gibbs R.A., Rubin G.M., Venter C.J.;
RL   Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
RN   [5]
RP   SEQUENCE FROM N.A.
RA   FlyBase;
RL   Submitted (SEP-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AE003643; AAN10875.1; -.
DR   FlyBase; FBgn0004002; wb.
DR   GO; GO:0007267; P:cell-cell signaling; NAS.
DR   InterPro; IPR001589; Actbind_actnin.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR006209; EGF_like.
DR   InterPro; IPR006210; IEGF.
DR   InterPro; IPR000034; Laminin_B.
DR   InterPro; IPR002049; Laminin_EGF.
DR   InterPro; IPR001791; Laminin_G.
DR   InterPro; IPR008211; LamNT.
DR   Pfam; PF00052; laminin_B; 2.
DR   Pfam; PF00053; laminin_EGF; 14.
DR   Pfam; PF00054; laminin_G; 4.
DR   Pfam; PF00055; laminin_Nterm; 1.
DR   PRINTS; PR00011; EGFLAMININ.
DR   SMART; SM00181; EGF; 9.
DR   SMART; SM00180; EGF_Lam; 17.
DR   SMART; SM00281; LamB; 2.
DR   SMART; SM00282; LamG; 5.
DR   SMART; SM00136; LamNT; 1.
DR   PROSITE; PS00019; ACTININ_1; 1.
DR   PROSITE; PS00022; EGF_1; 14.
DR   PROSITE; PS01186; EGF_2; 4.
DR   PROSITE; PS01248; LAMININ_TYPE_EGF; 16.
DR   PROSITE; PS50025; LAM_G_DOMAIN; 4.
SQ   SEQUENCE   3375 AA;  374742 MW;  38CB65C01BB6E416 CRC64;

  Query Match             3.2%;  Score 9;  DB 5;  Length 3375;
  Best Local Similarity   100.0%;  Pred. No. 11;
  Matches    9;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        182 NDGEWHRVT 190
              |||||||||
Db       3089 NDGEWHRVT 3097

RESULT 10
Q8BMI5
ID   Q8BMI5      PRELIMINARY;      PRT;   219 AA.
AC   Q8BMI5;
DT   01-MAR-2003 (TrEMBLrel. 23, Created)
DT   01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Weakly similar to neurogenic locus notch 3 protein (Fragment).
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6J; TISSUE=Forelimb;
RX   MEDLINE=22354683; PubMed=12466851;
RA   The FANTOM Consortium,
RA   the RIKEN Genome Exploration Research Group Phase I & II Team;
RT   "Analysis of the mouse transcriptome based on functional annotation of
RT   60,770 full-length cDNAs.";
RL   Nature 420:563-573(2002).
DR   EMBL; AK031068; BAC27237.1; -.
DR   PIR; PT0633; PT0633.
DR   GO; GO:0005509; F:calcium ion binding; IEA.
DR   InterPro; IPR000152; Asx_hydroxyl_S.
DR   InterPro; IPR000742; EGF_2.
DR   InterPro; IPR001881; EGF_Ca.
DR   InterPro; IPR006209; EGF_like.
DR   InterPro; IPR006210; IEGF.
DR   Pfam; PF00008; EGF; 2.
DR   SMART; SM00181; EGF; 3.
DR   SMART; SM00179; EGF_CA; 2.
DR   PROSITE; PS00010; ASX_HYDROXYL; 1.
DR   PROSITE; PS00022; EGF_1; 4.
DR   PROSITE; PS01186; EGF_2; 2.
FT   NON_TER       1      1
SQ   SEQUENCE   219 AA;  23098 MW;  9F2C4BE237F9B522 CRC64;

  Query Match             2.8%;  Score 8;  DB 11;  Length 219;
  Best Local Similarity   100.0%;  Pred. No. 11;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         61 PGFSGPRC 68
              ||||||||
Db         56 PGFSGPRC 63

RESULT 11
Q81XS5
ID   Q81XS5      PRELIMINARY;      PRT;   494 AA.
AC   Q81XS5;
DT   01-JUN-2003 (TrEMBLrel. 24, Created)
DT   01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Cytosol aminopeptidase.
GN   PEPA OR BA5155.
OS   Bacillus anthracis (strain Ames).
OC   Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX   NCBI_TaxID=198094;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=22608414; PubMed=12721629;
RA   Read T.D., Peterson S.N., Tourasse N., Baillie L.W., Paulsen I.T.,
RA   Nelson K.E., Tettelin H., Fouts D.E., Eisen J.A., Gill S.R.,
RA   Holtzapple E.K., Okstad O.A., Helgason E., Rilstone J., Wu M.,
RA   Kolonay J.F., Beanan M.J., Dodson R.J., Brinkac L.M., Gwinn M.,
RA   DeBoy R.T., Madpu R., Daugherty S.C., Durkin A.S., Haft D.H.,
RA   Nelson W.C., Peterson J.D., Pop M., Khouri H.M., Radune D.,
RA   Benton J.L., Mahamoud Y., Jiang L., Hance I.R., Weidman J.F.,
RA   Berry K.J., Plaut R.D., Wolf A.M., Watkins K.L., Nierman W.C.,
RA   Hazen A., Cline R., Redmond C., Thwaite J.E., White O., Salzberg S.L.,
RA   Thomason B., Friedlander A.M., Koehler T.M., Hanna P.C., Kolsto A.-B.,
RA   Fraser C.M.;
RT   "The genome sequence of Bacillus anthracis Ames and comparison to
RT   closely related bacteria.";
RL   Nature 423:81-86(2003).
DR   EMBL; AE017040; AAP28827.1; -.
DR   TIGR; BA5155; -.
DR   GO; GO:0005622; C:intracellular; IEA.
DR   GO; GO:0004177; F:aminopeptidase activity; IEA.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR   InterPro; IPR008330; Ctsl_amnpptdse_B.
DR   InterPro; IPR000819; Peptidase_M17_C.
```

```
DR   InterPro; IPR008283; Peptidase_M17_N.
DR   Pfam; PF00883; Peptidase_M17; 1.
DR   Pfam; PF02789; Peptidase_M17_N; 1.
DR   PRINTS; PR00481; LAMNOPPTDASE.
DR   PROSITE; PS00631; CYTOSOL_AP; 1.
DR   PIRSF; PIRSF036388; Ctsl_amnpptdse_B; 1.
KW   Aminopeptidase; Complete proteome.
SQ   SEQUENCE   494 AA;  53513 MW;  D400D6CC81FA31DF CRC64;

  Query Match             2.8%;  Score 8;  DB 16;  Length 494;
  Best Local Similarity   100.0%;  Pred. No. 22;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         232 DVATLTGG 239
               ||||||||
Db         366 DVATLTGG 373

RESULT 12
Q816E3
ID   Q816E3      PRELIMINARY;      PRT;   494 AA.
AC   Q816E3;
DT   01-JUN-2003 (TrEMBLrel. 24, Created)
DT   01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Cytosol aminopeptidase (EC 3.4.11.1).
GN   BC4921.
OS   Bacillus cereus (strain ATCC 14579 / DSM 31).
OC   Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX   NCBI_TaxID=226900;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=22608415; PubMed=12721630;
RA   Ivanova N., Sorokin A., Anderson I., Galleron N., Candelon B.,
RA   Kapatral V., Bhattacharyya A., Reznik G., Mikhailova N., Lapidus A.,
RA   Chu L., Mazur M., Goltsman E., Larsen N., D'Souza M., Walunas T.,
RA   Grechkin Y., Pusch G., Haselkorn R., Fonstein M., Ehrlich S.D.,
RA   Overbeek R., Kyrpides N.;
RT   "Genome sequence of Bacillus cereus and comparative analysis with
RT   Bacillus anthracis.";
RL   Nature 423:87-91(2003).
DR   EMBL; AE017013; AAP11794.1; -.
DR   GO; GO:0005622; C:intracellular; IEA.
DR   GO; GO:0016787; F:hydrolase activity; IEA.
DR   GO; GO:0004178; F:leucyl aminopeptidase activity; IEA.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR   InterPro; IPR008330; Ctsl_amnpptdse_B.
DR   InterPro; IPR000819; Peptidase_M17_C.
DR   InterPro; IPR008283; Peptidase_M17_N.
DR   Pfam; PF00883; Peptidase_M17; 1.
DR   Pfam; PF02789; Peptidase_M17_N; 1.
DR   PRINTS; PR00481; LAMNOPPTDASE.
DR   PROSITE; PS00631; CYTOSOL_AP; 1.
DR   PIRSF; PIRSF036388; Ctsl_amnpptdse_B; 1.
KW   Aminopeptidase; Hydrolase; Complete proteome.
SQ   SEQUENCE   494 AA;  53617 MW;  09E371EC83B0FBDF CRC64;

  Query Match             2.8%;  Score 8;  DB 16;  Length 494;
  Best Local Similarity   100.0%;  Pred. No. 22;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         232 DVATLTGG 239
               ||||||||
Db         366 DVATLTGG 373

RESULT 13
Q7X2C8
ID   Q7X2C8      PRELIMINARY;      PRT;   497 AA.
AC   Q7X2C8;
DT   01-OCT-2003 (TrEMBLrel. 25, Created)
DT   01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
```

```
DE   Thermostable leucine aminopeptidase.
GN   LAP.
OS   Geobacillus kaustophilus.
OC   Bacteria; Firmicutes; Bacillales; Bacillaceae; Geobacillus.
OX   NCBI_TaxID=1462;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Lin L.-L., Hsu W.-H.;
RT   "Bacillus kaustophilus thermostable leucine aminopeptidase.";
RL   Submitted (MAY-2003) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AY308074; AAP73418.1; -.
KW   Aminopeptidase.
SQ   SEQUENCE   497 AA;  53785 MW;  2793B98222DBE695 CRC64;

  Query Match             2.8%;  Score 8;  DB 2;  Length 497;
  Best Local Similarity   100.0%;  Pred. No. 22;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         232 DVATLTGG 239
               ||||||||
Db         371 DVATLTGG 378

RESULT 14
Q8YWD5
ID   Q8YWD5      PRELIMINARY;      PRT;   497 AA.
AC   Q8YWD5;
DT   01-MAR-2002 (TrEMBLrel. 20, Created)
DT   01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT   01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE   Sodium/solute symporter.
GN   ALL1678.
OS   Anabaena sp. (strain PCC 7120).
OC   Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Nostoc.
OX   NCBI_TaxID=103690;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=21595285; PubMed=11759840;
RA   Kaneko T., Nakamura Y., Wolk C.P., Kuritz T., Sasamoto S.,
RA   Watanabe A., Iriguchi M., Ishikawa A., Kawashima K., Kimura T.,
RA   Kishida Y., Kohara M., Matsumoto M., Matsuno A., Muraki A.,
RA   Nakazaki N., Shimpo S., Sugimoto M., Takazawa M., Yamada M.,
RA   Yasuda M., Tabata S.;
RT   "Complete genomic sequence of the filamentous nitrogen-fixing
RT   cyanobacterium Anabaena sp. strain PCC 7120.";
RL   DNA Res. 8:205-213(2001).
DR   EMBL; AP003586; BAB78044.1; -.
DR   PIR; AH2015; AH2015.
DR   GO; GO:0016020; C:membrane; IEA.
DR   GO; GO:0005215; F:transporter activity; IEA.
DR   GO; GO:0006810; P:transport; IEA.
DR   InterPro; IPR001734; Na/solut_symport.
DR   Pfam; PF00474; SSF; 1.
DR   PROSITE; PS50283; NA_SOLUT_SYMP_3; 1.
KW   Complete proteome.
SQ   SEQUENCE   497 AA;  53784 MW;  DF7B1111618FE222 CRC64;

  Query Match             2.8%;  Score 8;  DB 16;  Length 497;
  Best Local Similarity   100.0%;  Pred. No. 22;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         130 STASGLLL 137
               ||||||||
Db         325 STASGLLL 332

RESULT 15
O57409
ID   O57409      PRELIMINARY;      PRT;   615 AA.
AC   O57409;
DT   01-JUN-1998 (TrEMBLrel. 06, Created)
DT   01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
```

```
DE   DeltaB.
GN   DLB OR DELTAB.
OS   Brachydanio rerio (Zebrafish) (Danio rerio).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
OC   Cyprinidae; Danio.
OX   NCBI_TaxID=7955;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=98165391; PubMed=9425132;
RA   Haddon C., Smithers L., Schneider-Maunoury S., Coche T., Henrique D.,
RA   Lewis J.;
RT   "Multiple delta genes and lateral inhibition in zebrafish primary
RT   neurogenesis.";
RL   Development 125:359-370(1998).
DR   EMBL; AF006488; AAC41241.1; -.
DR   HSSP; P00740; 1EDM.
DR   ZFIN; ZDB-GENE-980526-114; dlb.
DR   GO; GO:0016020; C:membrane; IEA.
DR   GO; GO:0005509; F:calcium ion binding; IEA.
DR   GO; GO:0005198; F:structural molecule activity; IEA.
DR   GO; GO:0007154; P:cell communication; IEA.
DR   InterPro; IPR000152; Asx_hydroxyl_S.
DR   InterPro; IPR001774; DSL.
DR   InterPro; IPR000742; EGF_2.
DR   InterPro; IPR001881; EGF_Ca.
DR   InterPro; IPR001438; EGF_II.
DR   InterPro; IPR006209; EGF_like.
DR   InterPro; IPR002049; Laminin_EGF.
DR   Pfam; PF01414; DSL; 1.
DR   Pfam; PF00008; EGF; 7.
DR   PRINTS; PR00010; EGFBLOOD.
DR   PRINTS; PR00011; EGFLAMININ.
DR   SMART; SM00051; DSL; 1.
DR   SMART; SM00179; EGF_CA; 3.
DR   PROSITE; PS00010; ASX_HYDROXYL; 2.
DR   PROSITE; PS00022; EGF_1; 9.
DR   PROSITE; PS01186; EGF_2; 7.
DR   PROSITE; PS01187; EGF_CA; 2.
KW   EGF-like domain.
SQ   SEQUENCE   615 AA;  67592 MW;  CA18004428F5603C CRC64;

  Query Match             2.8%;  Score 8;  DB 13;  Length 615;
  Best Local Similarity   100.0%;  Pred. No. 27;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         45 PCLHGGTC 52
              ||||||||
Db        471 PCLHGGTC 478

Search completed: March  9, 2004, 17:30:34
Job time : 25.8707 secs
```

```
                    GenCore version 5.1.6
           Copyright (c) 1993 - 2004  Compugen Ltd.

OM protein - protein search, using sw model

Run on:          March  9, 2004, 17:11:22 ; Search time 25.5713 Seconds
                                            (without alignments)
                                            2320.373 Million cell updates/sec

Title:           US-10-006-011A-10
Perfect score:   1115
Sequence:        1 GIAESDWHLEGSGGNDAPGQ.........QPLDLQHRAQAGANTRPCPS 210

Scoring table:   BLOSUM62
                 Gapop 10.0 , Gapext 0.5

Searched:        1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters:      1586107

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :        A_Geneseq_29Jan04:*
                 1:  geneseqp1980s:*
                 2:  geneseqp1990s:*
                 3:  geneseqp2000s:*
                 4:  geneseqp2001s:*
                 5:  geneseqp2002s:*
                 6:  geneseqp2003as:*
                 7:  geneseqp2003bs:*
                 8:  geneseqp2004s:*

      Pred. No. is the number of results predicted by chance to have a
      score greater than or equal to the score of the result being printed,
      and is derived by analysis of the total score distribution.

                              SUMMARIES

                      %
Result               Query
  No.     Score     Match  Length  DB  ID                           Description
-------------------------------------------------------------------------------
     1     1115     100.0    4391   6  AAE34390                     Aae34390 Human per
     2     1114      99.9    4393   4  AAB31889                     Aab31889 Amino aci
     3     1092      97.9    4436   4  ABG23265                     Abg23265 Novel hum
     4     1030      92.4     195   4  AAB31890                     Aab31890 Amino aci
     5      292      26.2    1298   4  ABB61948                     Abb61948 Drosophil
     6    279.5      25.1     152   5  ABP56528                     Abp56528 BM hepara
     7    278.5      25.0     152   7  ADB65184                     Adb65184 Human pro
     8    278.5      25.0     210   4  AAU19782                     Aau19782 Human nov
     9    278.5      25.0     210   5  ABP48002                     Abp48002 Human pol
    10    278.5      25.0     210   7  ADC10964                     Adc10964 Human ext
    11    278.5      25.0     238   4  AAU18145                     Aau18145 Novel hum
    12    278.5      25.0     238   4  AAU17011                     Aau17011 Human nov
    13    278.5      25.0     238   4  ABB10450                     Abb10450 Human cDN
    14    278.5      25.0     238   4  AAU19961                     Aau19961 Novel hum
    15    278.5      25.0     238   5  ABJ05772                     Abj05772 Novel hum
    16    278.5      25.0     238   5  ABP67037                     Abp67037 Human pol
    17    278.5      25.0     375   4  AAU07421                     Aau07421 Novel hum
    18    278.5      25.0     432   4  AAU18102                     Aau18102 Novel hum
    19    278.5      25.0     432   4  AAU16938                     Aau16938 Human nov
    20    278.5      25.0     432   4  ABB10233                     Abb10233 Human cDN
    21    278.5      25.0     432   4  AAU19905                     Aau19905 Novel hum
    22    278.5      25.0     432   5  ABJ05729                     Abj05729 Novel hum
    23    278.5      25.0     432   5  ABP66820                     Abp66820 Human pol
    24    278.5      25.0     463   7  ADB64904                     Adb64904 Human pro
    25    278.5      25.0     671   6  ABP58231                     Abp58231 Human cel
```

```
    26    278.5      25.0    1009   7  ADE28105                     Ade28105 Human NTR
    27    275.5      24.7     819   5  ABB72291                     Abb72291 Rat prote
    28      265      23.8     406   5  ABP41801                     Abp41801 Human ova
    29      265      23.8     416   2  AAY73993                     Aay73993 Human pro
    30      265      23.8    1544   7  ADC39162                     Adc39162 Novel hum
    31      265      23.8    1565   7  ADC39156                     Adc39156 Novel hum
    32      265      23.8    1566   7  ADC39166                     Adc39166 Novel hum
    33      265      23.8    1931   6  ABU52400                     Abu52400 Human GPC
    34      265      23.8    2053   7  ADC39154                     Adc39154 Novel hum
    35      261      23.4     492   2  AAW26609                     Aaw26609 Human agr
    36    253.5      22.7    1741   5  ABP43859                     Abp43859 Human mRN
    37    253.5      22.7    2143   7  ADC39164                     Adc39164 Novel hum
    38      240      21.5     194   4  AAU19887                     Aau19887 Human nov
    39      240      21.5     194   5  ABP48107                     Abp48107 Human pol
    40      240      21.5     194   7  ADC11069                     Adc11069 Human pro
    41      226      20.3    4072   4  ABB63614                     Abb63614 Drosophil
    42    223.5      20.0    1130   2  AAR13436                     Aar13436 Merosin M
    43    223.5      20.0    1130   2  AAR71729                     Aar71729 Merosin m
    44    223.5      20.0    3070   5  AAO17359                     Aao17359 Human lam
    45    223.5      20.0    3088   3  AAB19794                     Aab19794 Human lam

                              ALIGNMENTS


RESULT 1
AAE34390
ID   AAE34390 standard; protein; 4391 AA.
XX
AC   AAE34390;
XX
DT   14-MAY-2003  (first entry)
XX
DE   Human perlecan protein.
XX
KW   Human; diagnosis; osteoarthritis; rheumatoid arthritis; perlecan.
XX
OS   Homo sapiens.
XX
PN   WO200295415-A2.
XX
PD   28-NOV-2002.
XX
PF   22-MAY-2002; 2002WO-EP005612.
XX
PR   23-MAY-2001; 2001GB-00012626.
XX
PA   (OSTE-) OSTEOMETER BIO TECH AS.
XX
PI   Christgau S,  Henriksen DB,  Cloos PAC;
XX
DR   WPI; 2003-140389/13.
XX
PT   An assay for the diagnosis or assessment of the severity of
PT   osteoarthritis or rheumatoid arthritis comprising detecting an isomerized
PT   or optically inverted protein in a sample.
XX
PS   Disclosure; Page 46-67; 106pp; English.
XX
CC   The invention relates to an assay for the diagnosis or assessment of the
CC   severity of osteoarthritis or rheumatoid arthritis. The assay involves
CC   measuring (in a biological sample) the amount or presence of an
CC   isomerised or optically inverted protein or one or more isomerised or
CC   optically inverted fragments from proteins such as perlecan, biglycan,
CC   decorin, fibrillin-1 or protocadherin. The assay is useful for the
CC   diagnosis or assessment of the severity of osteoarthritis or rheumatoid
CC   arthritis. The present sequence is human perlecan protein
XX
SQ   Sequence 4391 AA;

  Query Match             100.0%;  Score 1115;  DB 6;  Length 4391;
  Best Local Similarity   100.0%;  Pred. No. 1.2e-98;
  Matches  210;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;
```

```
Qy          1 GIAESDWHLEGSGGNDAPGQYGAYFHDDGFLAFPGHVFSRSLPEVPETIELEVRTSTASG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4182 GIAESDWHLEGSGGNDAPGQYGAYFHDDGFLAFPGHVFSRSLPEVPETIELEVRTSTASG 4241

Qy         61 LLLWQGVEVGEAGQGKDFISLGLQDGHLVFRYQLGSGEARLVSEDPINDGEWHRVTALRE 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4242 LLLWQGVEVGEAGQGKDFISLGLQDGHLVFRYQLGSGEARLVSEDPINDGEWHRVTALRE 4301

Qy        121 GRRGSIQVDGEELVSGRSPGPNVAVNAKGSVYIGGAPDVATLTGGRFSSGITGCVKNLVL 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4302 GRRGSIQVDGEELVSGRSPGPNVAVNAKGSVYIGGAPDVATLTGGRFSSGITGCVKNLVL 4361

Qy        181 HSARPGAPPPQPLDLQHRAQAGANTRPCPS 210
              ||||||||||||||||||||||||||||||
Db       4362 HSARPGAPPPQPLDLQHRAQAGANTRPCPS 4391

RESULT 2
AAB31889
ID   AAB31889 standard; protein; 4393 AA.
XX
AC   AAB31889;
XX
DT   15-MAY-2001  (first entry)
XX
DE   Amino acid sequence of a human protein.
XX
KW   Human; perlecan; retinol-binding plasma protein; calgranulin B; vaccine;
KW   ganglioside GM2 activator; saposin B; degenerative disease; glial cell;
KW   neurological disease; auto-immune disease; multiple sclerosis; toxicity;
KW   Alzheimer's disease; Parkinson's disease; amyotrophic lateral sclerosis;
KW   rheumatoid polyarthritis; lupus erythematosus; gene therapy.
XX
OS   Homo sapiens.
XX
PN   WO200105422-A2.
XX
PD   25-JAN-2001.
XX
PF   17-JUL-2000; 2000WO-FR002057.
XX
PR   15-JUL-1999;   99FR-00009372.
XX
PA   (INMR ) BIOMERIEUX STELHYS.
XX
PI   Roecklin D,  Kolbe H,  Charles M,  Malcus C,  Santoro L,  Perron H;
XX
DR   WPI; 2001-159475/16.
XX
PT   Detecting, preventing and treating degenerative, neurological and
PT   autoimmune diseases, particularly multiple sclerosis, using specified
PT   polypeptides or related nucleic acid or ligand.
XX
PS   Claim 1; Page 138-152; 209pp; French.
XX
CC   The present sequence represents a human protein, which is used in the
CC   method of the invention. The specification describes a method which uses
CC   at least one polypeptide or polynucleotide sequence belonging to the
CC   perlecan, precursor of the retinol-binding plasma protein, precursor of
CC   the ganglioside GM2 activator, calgranulin B or saposin B protein
CC   families. The method is used for detecting, preventing or treating a
CC   degenerative, neurological and/or auto-immune disease. The
CC   polynucleotides and polypeptides are used for diagnosis, prognosis,
CC   prevention and treatment of multiple sclerosis (in its various forms and
CC   phases). They may also be useful in cases of e.g. Alzheimer's and
CC   Parkinson's diseases, amyotrophic lateral sclerosis, rheumatoid
CC   polyarthritis and lupus erythematosus, including use as vaccines and in
CC   gene therapy (expression of sense or antisense sequences). They can also
CC   be used to assess efficacy of potential therapeutic agents, particularly
CC   compounds that reduce or inhibit toxicity towards glial cells
XX
SQ   Sequence 4393 AA;
```

```
  Query Match             99.9%;  Score 1114;  DB 4;  Length 4393;
  Best Local Similarity   99.5%;  Pred. No. 1.5e-98;
  Matches  209;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 GIAESDWHLEGSGGNDAPGQYGAYFHDDGFLAFPGHVFSRSLPEVPETIELEVRTSTASG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4184 GIAESDWHLEGSGGNDAPGQYGAYFHDDGFLAFPGHVFSRSLPEVPETIELEVRTSTASG 4243

Qy         61 LLLWQGVEVGEAGQGKDFISLGLQDGHLVFRYQLGSGEARLVSEDPINDGEWHRVTALRE 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4244 LLLWQGVEVGEAGQGKDFISLGLQDGHLVFRYQLGSGEARLVSEDPINDGEWHRVTALRE 4303

Qy        121 GRRGSIQVDGEELVSGRSPGPNVAVNAKGSVYIGGAPDVATLTGGRFSSGITGCVKNLVL 180
              ||||||||||||||||||||||||||||||:|||||||||||||||||||||||||||||
Db       4304 GRRGSIQVDGEELVSGRSPGPNVAVNAKGSIYIGGAPDVATLTGGRFSSGITGCVKNLVL 4363

Qy        181 HSARPGAPPPQPLDLQHRAQAGANTRPCPS 210
              ||||||||||||||||||||||||||||||
Db       4364 HSARPGAPPPQPLDLQHRAQAGANTRPCPS 4393

RESULT 3
ABG23265
ID   ABG23265 standard; protein; 4436 AA.
XX
AC   ABG23265;
XX
DT   18-FEB-2002  (first entry)
XX
DE   Novel human diagnostic protein #23256.
XX
KW   Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW   food supplement; medical imaging; diagnostic; genetic disorder.
XX
OS   Homo sapiens.
XX
PN   WO200175067-A2.
XX
PD   11-OCT-2001.
XX
PF   30-MAR-2001; 2001WO-US008631.
XX
PR   31-MAR-2000; 2000US-00540217.
PR   23-AUG-2000; 2000US-00649167.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Drmanac RT,  Liu C,  Tang YT;
XX
DR   WPI; 2001-639362/73.
DR   N-PSDB; AAS87452.
XX
PT   New isolated polynucleotide and encoded polypeptides, useful in
PT   diagnostics, forensics, gene mapping, identification of mutations
PT   responsible for genetic disorders or other traits and to assess
PT   biodiversity.
XX
PS   Claim 20; SEQ ID NO 53624; 103pp; English.
XX
CC   The invention relates to isolated polynucleotide (I) and polypeptide (II)
CC   sequences. (I) is useful as hybridisation probes, polymerase chain
CC   reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
CC   and in recombinant production of (II). The polynucleotides are also used
CC   in diagnostics as expressed sequence tags for identifying expressed
CC   genes. (I) is useful in gene therapy techniques to restore normal
CC   activity of (II) or to treat disease states involving (II). (II) is
CC   useful for generating antibodies against it, detecting or quantitating a
CC   polypeptide in tissue, as molecular weight markers and as a food
CC   supplement. (II) and its binding partners are useful in medical imaging
CC   of sites expressing (II). (I) and (II) are useful for treating disorders
CC   involving aberrant protein expression or biological activity. The
CC   polypeptide and polynucleotide sequences have applications in
```

```
CC   diagnostics, forensics, gene mapping, identification of mutations
CC   responsible for genetic disorders or other traits to assess biodiversity
CC   and to produce other types of data and products dependent on DNA and
CC   amino acid sequences. ABG00010-ABG30377 represent novel human diagnostic
CC   amino acid sequences of the invention. Note: The sequence data for this
CC   patent did not appear in the printed specification, but was obtained in
CC   electronic format directly from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 4436 AA;

  Query Match             97.9%;  Score 1092;  DB 4;  Length 4436;
  Best Local Similarity   95.0%;  Pred. No. 2.1e-96;
  Matches  208;  Conservative    0;  Mismatches    1;  Indels   10;  Gaps    1;

Qy          1 GIAESDWHLEGSGGNDAPGQYGAYFHDDGFLAFPGHVFSRSLPEVPETIELEVRTSTASG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4217 GIAESDWHLEGSGGNDAPGQYGAYFHDDGFLAFPGHVFSRSLPEVPETIELEVRTSTASG 4276

Qy         61 LLLWQGVEVGEAGQGKDFISLGLQDGHLVF----------RYQLGSGEARLVSEDPINDG 110
              ||||||||||||||||||||||||||||||          |||||||||||||||| |||
Db       4277 LLLWQGVEVGEAGQGKDFISLGLQDGHLVFSSYFLGLDDCRYQLGSGEARLVSEDPFNDG 4336

Qy        111 EWHRVTALREGRRGSIQVDGEELVSGRSPGPNVAVNAKGSVYIGGAPDVATLTGGRFSSG 170
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4337 EWHRVTALREGRRGSIQVDGEELVSGRSPGPNVAVNAKGSVYIGGAPDVATLTGGRFSSG 4396

Qy        171 ITGCVKNLVLHSARPGAPPPQPLDLQHRAQAGANTRPCP 209
              |||||||||||||||||||||||||||||||||||||||
Db       4397 ITGCVKNLVLHSARPGAPPPQPLDLQHRAQAGANTRPCP 4435


RESULT 4
AAB31890
ID   AAB31890 standard; protein; 195 AA.
XX
AC   AAB31890;
XX
DT   15-MAY-2001  (first entry)
XX
DE   Amino acid sequence of the C-terminal of the human perlecan protein.
XX
KW   Human; perlecan; retinol-binding plasma protein; calgranulin B; vaccine;
KW   ganglioside GM2 activator; saposin B; degenerative disease; glial cell;
KW   neurological disease; auto-immune disease; multiple sclerosis; toxicity;
KW   Alzheimer's disease; Parkinson's disease; amyotrophic lateral sclerosis;
KW   rheumatoid polyarthritis; lupus erythematosus; gene therapy.
XX
OS   Homo sapiens.
XX
PN   WO200105422-A2.
XX
PD   25-JAN-2001.
XX
PF   17-JUL-2000; 2000WO-FR002057.
XX
PR   15-JUL-1999;   99FR-00009372.
XX
PA   (INMR ) BIOMERIEUX STELHYS.
XX
PI   Roecklin D,  Kolbe H,  Charles M,  Malcus C,  Santoro L,  Perron H;
XX
DR   WPI; 2001-159475/16.
DR   N-PSDB; AAF54728.
XX
PT   Detecting, preventing and treating degenerative, neurological and
PT   autoimmune diseases, particularly multiple sclerosis, using specified
PT   polypeptides or related nucleic acid or ligand.
XX
PS   Claim 1; Page 152-153; 209pp; French.
XX
CC   The present sequence represents a human polypeptide, which is used in the
CC   method of the invention. The specification describes a method which uses
```

```
CC   at least one polypeptide or polynucleotide sequence belonging to the
CC   perlecan, precursor of the retinol-binding plasma protein, precursor of
CC   the ganglioside GM2 activator, calgranulin B or saposin B protein
CC   families. The method is used for detecting, preventing or treating a
CC   degenerative, neurological and/or auto-immune disease. The
CC   polynucleotides and polypeptides are used for diagnosis, prognosis,
CC   prevention and treatment of multiple sclerosis (in its various forms and
CC   phases). They may also be useful in cases of e.g. Alzheimer's and
CC   Parkinson's diseases, amyotrophic lateral sclerosis, rheumatoid
CC   polyarthritis and lupus erythematosus, including use as vaccines and in
CC   gene therapy (expression of sense or antisense sequences). They can also
CC   be used to assess efficacy of potential therapeutic agents, particularly
CC   compounds that reduce or inhibit toxicity towards glial cells
XX
SQ   Sequence 195 AA;

  Query Match             92.4%;  Score 1030;  DB 4;  Length 195;
  Best Local Similarity   100.0%;  Pred. No. 3.5e-92;
  Matches  195;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         16 DAPGQYGAYFHDDGFLAFPGHVFSRSLPEVPETIELEVRTSTASGLLLWQGVEVGEAGQG 75
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 DAPGQYGAYFHDDGFLAFPGHVFSRSLPEVPETIELEVRTSTASGLLLWQGVEVGEAGQG 60

Qy         76 KDFISLGLQDGHLVFRYQLGSGEARLVSEDPINDGEWHRVTALREGRRGSIQVDGEELVS 135
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 KDFISLGLQDGHLVFRYQLGSGEARLVSEDPINDGEWHRVTALREGRRGSIQVDGEELVS 120

Qy        136 GRSPGPNVAVNAKGSVYIGGAPDVATLTGGRFSSGITGCVKNLVLHSARPGAPPPQPLDL 195
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GRSPGPNVAVNAKGSVYIGGAPDVATLTGGRFSSGITGCVKNLVLHSARPGAPPPQPLDL 180

Qy        196 QHRAQAGANTRPCPS 210
              |||||||||||||||
Db        181 QHRAQAGANTRPCPS 195


RESULT 5
ABB61948
ID   ABB61948 standard; protein; 1298 AA.
XX
AC   ABB61948;
XX
DT   26-MAR-2002  (first entry)
XX
DE   Drosophila melanogaster polypeptide SEQ ID NO 12636.
XX
KW   Drosophila; developmental biology; cell signalling; insecticide;
KW   pharmaceutical.
XX
OS   Drosophila melanogaster.
XX
PN   WO200171042-A2.
XX
PD   27-SEP-2001.
XX
PF   23-MAR-2001; 2001WO-US009231.
XX
PR   23-MAR-2000; 2000US-0191637P.
PR   11-JUL-2000; 2000US-00614150.
XX
PA   (PEKE ) PE CORP NY.
XX
PI   Venter JC,  Adams M,  Li PWD,  Myers EW;
XX
DR   WPI; 2001-656860/75.
DR   N-PSDB; ABL06051.
XX
PT   New isolated nucleic acid detection reagent for detecting 1000 or more
PT   genes from Drosophila and for elucidating cell signaling and cell-cell
PT   interactions.
XX
PS   Disclosure; SEQ ID NO 12636; 21pp + Sequence Listing; English.
```

```
XX
CC   The invention relates to an isolated nucleic acid detection reagent
CC   capable of detecting 1000 or more genes from Drosophila. The invention is
CC   useful in developmental biology and in elucidating cell signalling and
CC   cell-cell interactions in higher eukaryotes for the development of
CC   insecticides, therapeutics and pharmaceutical drugs. The invention
CC   discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC   sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-
CC   ABB72072). The sequence data for this patent did not form part of the
CC   printed specification, but was obtained in electronic format directly
CC   from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 1298 AA;

  Query Match             26.2%;  Score 292;  DB 4;  Length 1298;
  Best Local Similarity   35.0%;  Pred. No. 4.9e-19;
  Matches   64;  Conservative   37;  Mismatches   70;  Indels   12;  Gaps    5;

Qy         23 AYFH--DDGFLAFPGHVFSRSLPEVPETIELEVRTSTASGLLLWQGVEVGEAGQGKDFIS 80
              :|||  |   ::    | | |:      : | ::| : :|::|| | : |   :  |::|
Db       1123 SYFHYNDADTMS---QVISYSI-----DLNLRIKTHSENGVILWTGRQ-GTTEEHDDYLS 1173

Qy         81 LGLQDGHLVFRYQLGSGEARL-VSEDPINDGEWHRVTALREGRRGSIQVDGEELVSGRSP 139
              ||:: |:| ||| |||||  :  :   ::|| |||| |:|  : | ::||| : |: |:|
Db       1174 LGIEQGYLHFRYDLGSGEVDIRFNGTKVSDGLWHRVRAIRNSQEGYLEVDGRKTVTLRAP 1233

Qy        140 GPNVAVNAKGSVYIGGAPDVATLTGGRFSSGITGCVKNLVLHSARPGAPPPQPLDLQHRA 199
              |    :|    :|:|| |||   |  |: ||| ||:  :||         |  |  :|
Db       1234 GKLRQLNTDTGLYVGGMPDVGYFTHQRYFSGIVGCISEIVLAGEMKLNFDPNTLGTEHNV 1293

Qy        200 QAG 202
              : |
Db       1294 ETG 1296


RESULT 6
ABP56528
ID   ABP56528 standard; protein; 152 AA.
XX
AC   ABP56528;
XX
DT   21-MAR-2003  (first entry)
XX
DE   BM heparan sulfate protein glycan core protein 16.72 SEQ ID NO:2.
XX
KW   Basement membrane heparan sulfate protein glycan core protein 16.72;
KW   human; nerve system developmental deformity; paralysis; arrhythmia;
KW   bronchial asthma; peptic ulcer; dementia.
XX
OS   Homo sapiens.
XX
PN   CN1352029-A.
XX
PD   05-JUN-2002.
XX
PF   10-NOV-2000; 2000CN-00127356.
XX
PR   10-NOV-2000; 2000CN-00127356.
XX
PA   (BODE-) BODE GENE DEV CO LTD SHANGHAI.
XX
PI   Mao Y,  Xie Y;
XX
DR   WPI; 2002-714417/78.
DR   N-PSDB; ABZ22351.
XX
PT   New polypeptide-basement membrane heparan sulfate protein glycan core
PT   protein 16.72 and polynucleotide for encoding such polypeptide.
XX
PS   Claim 1; Page 27 (Disclosure); 33pp; Chinese.
XX
CC   The present sequence represents human basement membrane heparan sulfate
CC   protein glycan core protein 16.72 (I). Also described is a DNA
```

```
CC   recombination process used to produce (I). (I) can be used for treating
CC   various diseases, such as nerve system developmental deformity,
CC   paralysis, arrhythmia, bronchial asthma, peptic ulcer and dementia
XX
SQ   Sequence 152 AA;

  Query Match             25.1%;  Score 279.5;  DB 5;  Length 152;
  Best Local Similarity   37.3%;  Pred. No. 4.5e-19;
  Matches   59;  Conservative   25;  Mismatches   65;  Indels    9;  Gaps    2;

Qy         51 LEVRTSTASGLLLWQGVEVGEAGQGKDFISLGLQDGHLVFRYQLGSGEARLVSEDPINDG 110
              :  :|:   |||||:|          |||||||:|| ||| | |||| | ::     |||
Db          1 MRFKTTAKDGLLLWRG--DSPMRPNSDFISLGLRDGALVFSYNLGSGVASIMVNGSFNDG 58

Qy        111 EWHRVTALREGRRGSIQVDGEELVSGRSPGPNVAVNAKGSVYIGGAPDVATLTGGRFSSG 170
               |||| |:|:|: | | ||     :|:|||    :|  |::|:||  ::|  |  ::  |
Db         59 RWHRVKAVRDGQSGKITVDDYGARTGKSPGMMRQLNINGALYVGGMKEIALHTNRQYMRG 118

Qy        171 ITGCVKNLVLHSARPGAPPPQPLDLQHRAQAGANTRPC 208
              : ||: :  | :          : |   |  | |   |
Db        119 LVGCISHFTLST-------DYHISLAEDAVDGKNINTC 149


RESULT 7
ADB65184
ID   ADB65184 standard; protein; 152 AA.
XX
AC   ADB65184;
XX
DT   04-DEC-2003  (first entry)
XX
DE   Human protein encoded by clone SPLEN20183020.
XX
KW   Human; pharmaceutical; diagnostic; gene therapy; tissue regeneration;
KW   cell regeneration; membrane protein; signal transduction-related protein;
KW   transcription-related protein; osteoporosis; neurological disease;
KW   cancer; tumour.
XX
OS   Homo sapiens.
XX
PN   EP1308459-A2.
XX
PD   07-MAY-2003.
XX
PF   28-MAR-2002; 2002EP-00007401.
XX
PR   05-NOV-2001; 2001JP-00379298.
PR   25-JAN-2002; 2002US-00350978.
XX
PA   (HELI-) HELIX RES INST.
PA   (REAS-) RES ASSOC BIOTECHNOLOGY.
XX
PI   Isogai T,  Sugiyama T,  Otsuki T,  Wakamatsu A,  Sato H,  Ishii S;
PI   Yamamoto J,  Isono Y,  Hio Y,  Otsuka K,  Nagai K,  Irie R,  Tamechika I;
PI   Seki N,  Yoshikawa T,  Otsuka M,  Nagahari K,  Masuho Y;
XX
DR   WPI; 2003-450961/43.
DR   N-PSDB; ADB63214.
XX
PT   New polynucleotides and polypeptides, useful for developing a diagnostic
PT   marker or medicines for regulation of their expression and activity, or
PT   as targets of gene therapy.
XX
PS   Claim 1; Page; 222pp; English.
XX
CC   The invention discloses a polynucleotide comprising a sequence selected
CC   from 1970 fully defined nucleotide sequences which encode novel
CC   polypeptides. Also claimed is a polypeptide encoded by the polynucleotide
CC   or its partial peptide, an antibody binding to the polypeptide or peptide
CC   of the polynucleotide, immunologically assaying the polypeptide or
CC   peptide of the polynucleotide by contacting the polypeptide or peptide
CC   with the antibody of the encoded protein, and observing the binding
CC   between the two, a transformant carrying the polynucleotide in an
```

```
CC   expressible manner and an antisense polynucleotide. The oligonucleotide
CC   is useful as a primer for synthesising the polynucleotide, or as a probe
CC   for detecting the polynucleotide. The polynucleotides and encoded
CC   proteins are useful as pharmaceutical agents and many disease-related
CC   genes may be included in them, for developing a diagnostic marker or
CC   medicines for regulation of their expression and activity, or as targets
CC   of gene therapy. The genes are involved in tissue and/or cell
CC   regeneration. Membrane proteins, signal transduction-related proteins,
CC   transcription-related proteins, disease-related proteins and genes
CC   encoding them can be used as indicators for diseases (e.g. osteoporosis,
CC   neurological diseases, cancer, tumours. The cDNA may be used to regulate
CC   the activity or expression of the encoded protein to treat diseases. The
CC   sequence presented is a protein of the invention. Note: Some of the
CC   sequence data for this patent is not represented in the printed
CC   specification, but is based on sequence information supplied by the
CC   European Patent Office.
XX
SQ   Sequence 152 AA;

  Query Match             25.0%;  Score 278.5;  DB 7;  Length 152;
  Best Local Similarity   37.3%;  Pred. No. 5.7e-19;
  Matches   59;  Conservative   25;  Mismatches   65;  Indels    9;  Gaps    2;

Qy           51 LEVRTSTASGLLLWQGVEVGEAGQGKDFISLGLQDGHLVFRYQLGSGEARLVSEDPINDG 110
                :  :|:   |||||:|          |||||||:|| ||| | |||| | ::     |||
Db            1 MRFKTTAKDGLLLWRG--DSPMRPNSDFISLGLRDGALVFSYNLGSGVASIMVNGSFNDG 58

Qy          111 EWHRVTALREGRRGSIQVDGEELVSGRSPGPNVAVNAKGSVYIGGAPDVATLTGGRFSSG 170
                 |||| |:|:|: | | ||     :|:|||    :|  |::|:||  ::|  |  ::  |
Db           59 RWHRVKAVRDGQSGKITVDDYGARTGKSPGMMRQLNINGALYVGGMKEIALHTNRQYMRG 118

Qy          171 ITGCVKNLVLHSARPGAPPPQPLDLQHRAQAGANTRPC 208
                : ||: :  | :          :  |   |  | |   |
Db          119 LVGCISHFTLST-------DYHISLVEDAVDGKNINTC 149


RESULT 8
AAU19782
ID   AAU19782 standard; protein; 210 AA.
XX
AC   AAU19782;
XX
DT   04-DEC-2001  (first entry)
XX
DE   Human novel extracellular matrix protein, Seq ID No 432.
XX
KW   Human; secreted extracellular matrix protein; immunomodulatory; Anti-HIV;
KW   antianemic; antirheumatic; antisclerotic; cardiant; vascular;
KW   cerebroprotective; thrombolytic; antimicrobial; ophthalmic; cytostatic;
KW   antialzheimers; immune/autoimmune disease; HIV infection; anaemia;
KW   human immunodeficiency virus; rheumatoid arthritis; multiple sclerosis;
KW   cancers; hyperproliferative disorder; breast neoplasm; melanoma;
KW   Sezary syndrome; Gaucher's disease; neurological diseases;
KW   Alzheimer's disease; Parkinson's disease; cardiovascular disorder;
KW   cardiac arrest; tachycardia; angina; infection; corneal infections;
KW   wound healing; immunogen; gene therapy; antisense; food additive.
XX
OS   Homo sapiens.
XX
PN   WO200155368-A1.
XX
PD   02-AUG-2001.
XX
PF   17-JAN-2001; 2001WO-US001348.
XX
PR   31-JAN-2000; 2000US-0179065P.
PR   04-FEB-2000; 2000US-0180628P.
PR   24-FEB-2000; 2000US-0184664P.
PR   02-MAR-2000; 2000US-0186350P.
PR   16-MAR-2000; 2000US-0189874P.
PR   17-MAR-2000; 2000US-0190076P.
PR   18-APR-2000; 2000US-0198123P.
PR   19-MAY-2000; 2000US-0205515P.
```

```
PR   07-JUN-2000; 2000US-0209467P.
PR   28-JUN-2000; 2000US-0214886P.
PR   30-JUN-2000; 2000US-0215135P.
PR   07-JUL-2000; 2000US-0216647P.
PR   07-JUL-2000; 2000US-0216880P.
PR   11-JUL-2000; 2000US-0217487P.
PR   11-JUL-2000; 2000US-0217496P.
PR   14-JUL-2000; 2000US-0218290P.
PR   26-JUL-2000; 2000US-0220963P.
PR   26-JUL-2000; 2000US-0220964P.
PR   14-AUG-2000; 2000US-0224518P.
PR   14-AUG-2000; 2000US-0224519P.
PR   14-AUG-2000; 2000US-0225213P.
PR   14-AUG-2000; 2000US-0225214P.
PR   14-AUG-2000; 2000US-0225266P.
PR   14-AUG-2000; 2000US-0225267P.
PR   14-AUG-2000; 2000US-0225268P.
PR   14-AUG-2000; 2000US-0225270P.
PR   14-AUG-2000; 2000US-0225447P.
PR   14-AUG-2000; 2000US-0225757P.
PR   14-AUG-2000; 2000US-0225758P.
PR   14-AUG-2000; 2000US-0225759P.
PR   18-AUG-2000; 2000US-0226279P.
PR   22-AUG-2000; 2000US-0226681P.
PR   22-AUG-2000; 2000US-0226868P.
PR   22-AUG-2000; 2000US-0227182P.
PR   23-AUG-2000; 2000US-0227009P.
PR   30-AUG-2000; 2000US-0228924P.
PR   01-SEP-2000; 2000US-0229287P.
PR   01-SEP-2000; 2000US-0229343P.
PR   01-SEP-2000; 2000US-0229344P.
PR   01-SEP-2000; 2000US-0229345P.
PR   05-SEP-2000; 2000US-0229509P.
PR   05-SEP-2000; 2000US-0229513P.
PR   06-SEP-2000; 2000US-0230437P.
PR   06-SEP-2000; 2000US-0230438P.
PR   08-SEP-2000; 2000US-0231242P.
PR   08-SEP-2000; 2000US-0231243P.
PR   08-SEP-2000; 2000US-0231244P.
PR   08-SEP-2000; 2000US-0231413P.
PR   08-SEP-2000; 2000US-0231414P.
PR   08-SEP-2000; 2000US-0232080P.
PR   08-SEP-2000; 2000US-0232081P.
PR   12-SEP-2000; 2000US-0231968P.
PR   14-SEP-2000; 2000US-0232397P.
PR   14-SEP-2000; 2000US-0232398P.
PR   14-SEP-2000; 2000US-0232399P.
PR   14-SEP-2000; 2000US-0232400P.
PR   14-SEP-2000; 2000US-0232401P.
PR   14-SEP-2000; 2000US-0233063P.
PR   14-SEP-2000; 2000US-0233064P.
PR   14-SEP-2000; 2000US-0233065P.
PR   21-SEP-2000; 2000US-0234223P.
PR   21-SEP-2000; 2000US-0234274P.
PR   25-SEP-2000; 2000US-0234997P.
PR   25-SEP-2000; 2000US-0234998P.
PR   26-SEP-2000; 2000US-0235484P.
PR   27-SEP-2000; 2000US-0235834P.
PR   27-SEP-2000; 2000US-0235836P.
PR   29-SEP-2000; 2000US-0236327P.
PR   29-SEP-2000; 2000US-0236367P.
PR   29-SEP-2000; 2000US-0236368P.
PR   29-SEP-2000; 2000US-0236369P.
PR   29-SEP-2000; 2000US-0236370P.
PR   02-OCT-2000; 2000US-0236802P.
PR   02-OCT-2000; 2000US-0237037P.
PR   02-OCT-2000; 2000US-0237038P.
PR   02-OCT-2000; 2000US-0237039P.
PR   02-OCT-2000; 2000US-0237040P.
PR   13-OCT-2000; 2000US-0239935P.
PR   13-OCT-2000; 2000US-0239937P.
PR   20-OCT-2000; 2000US-0240960P.
PR   20-OCT-2000; 2000US-0241221P.
```

```
PR   20-OCT-2000; 2000US-0241785P.
PR   20-OCT-2000; 2000US-0241786P.
PR   20-OCT-2000; 2000US-0241787P.
PR   20-OCT-2000; 2000US-0241808P.
PR   20-OCT-2000; 2000US-0241809P.
PR   20-OCT-2000; 2000US-0241826P.
PR   01-NOV-2000; 2000US-0244617P.
PR   08-NOV-2000; 2000US-0246474P.
PR   08-NOV-2000; 2000US-0246475P.
PR   08-NOV-2000; 2000US-0246476P.
PR   08-NOV-2000; 2000US-0246477P.
PR   08-NOV-2000; 2000US-0246478P.
PR   08-NOV-2000; 2000US-0246523P.
PR   08-NOV-2000; 2000US-0246524P.
PR   08-NOV-2000; 2000US-0246525P.
PR   08-NOV-2000; 2000US-0246526P.
PR   08-NOV-2000; 2000US-0246527P.
PR   08-NOV-2000; 2000US-0246528P.
PR   08-NOV-2000; 2000US-0246532P.
PR   08-NOV-2000; 2000US-0246609P.
PR   08-NOV-2000; 2000US-0246610P.
PR   08-NOV-2000; 2000US-0246611P.
PR   08-NOV-2000; 2000US-0246613P.
PR   17-NOV-2000; 2000US-0249207P.
PR   17-NOV-2000; 2000US-0249208P.
PR   17-NOV-2000; 2000US-0249209P.
PR   17-NOV-2000; 2000US-0249210P.
PR   17-NOV-2000; 2000US-0249211P.
PR   17-NOV-2000; 2000US-0249212P.
PR   17-NOV-2000; 2000US-0249213P.
PR   17-NOV-2000; 2000US-0249214P.
PR   17-NOV-2000; 2000US-0249215P.
PR   17-NOV-2000; 2000US-0249216P.
PR   17-NOV-2000; 2000US-0249217P.
PR   17-NOV-2000; 2000US-0249218P.
PR   17-NOV-2000; 2000US-0249244P.
PR   17-NOV-2000; 2000US-0249245P.
PR   17-NOV-2000; 2000US-0249264P.
PR   17-NOV-2000; 2000US-0249265P.
PR   17-NOV-2000; 2000US-0249297P.
PR   17-NOV-2000; 2000US-0249299P.
PR   17-NOV-2000; 2000US-0249300P.
PR   01-DEC-2000; 2000US-0250160P.
PR   01-DEC-2000; 2000US-0250391P.
PR   05-DEC-2000; 2000US-0251030P.
PR   05-DEC-2000; 2000US-0251988P.
PR   05-DEC-2000; 2000US-0256719P.
PR   06-DEC-2000; 2000US-0251479P.
PR   08-DEC-2000; 2000US-0251856P.
PR   08-DEC-2000; 2000US-0251868P.
PR   08-DEC-2000; 2000US-0251869P.
PR   08-DEC-2000; 2000US-0251989P.
PR   08-DEC-2000; 2000US-0251990P.
PR   11-DEC-2000; 2000US-0254097P.
PR   05-JAN-2001; 2001US-0259678P.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Rosen CA,  Barash SC,  Ruben SM;
XX
DR   WPI; 2001-465572/50.
DR   N-PSDB; AAS31353.
XX
PT   Nucleic acid molecules encoding human secreted extracellular matrix
PT   proteins, used in preventing, treating or ameliorating a disorder, e.g.
PT   Alzheimer's and Parkinson's diseases and cancers.
XX
PS   Claim 11; SEQ ID NO 432; 577pp; English.
XX
CC   The invention relates to isolated nucleic acid molecules encoding novel
CC   human secreted extracellular matrix proteins (SPs). The polynucleotides
CC   and proteins are used to prevent, treat a medical condition in e.g.
CC   humans, mice, rabbits, goats, horses, cats, dogs, chickens or sheep. For
CC   example, disorders associated with decreased expression of SPs. The SP
CC   polynucleotide or a vector expressing them may be administered to treat
CC   diseases by gene therapy. Antisense molecules may be administered to down
CC   regulate expression of SPs by binding with the cells own genes and
CC   preventing their expression. The polynucleotides may also be used as DNA
CC   probes in diagnostic assays. The SPs may also be used as antigens to
CC   produce antibodies and to identify modulators (agonists and antagonists)
CC   of the SPs. The anti-(SP) antibodies and antagonists may also be used to
CC   down regulate expression and activity of SP and as diagnostic agents for
CC   detecting the presence of SPs in samples. The disorders include for
CC   example: immune/autoimmune diseases (e.g. HIV (human immunodeficiency
CC   virus) infections, anaemia, rheumatoid arthritis and multiple sclerosis),
CC   cancers and hyperproliferative disorders (e.g. melanomas, neoplasms of
CC   the breast or liver, Sezary syndrome and Gaucher's disease), neurological
CC   diseases (e.g. Alzheimer's disease, Parkinson's disease) cardio-
CC   /cerebrovascular disorders (e.g. cardiac arrest, tachycardia and angina),
CC   infections caused by bacteria, viruses and fungi and ocular disorders
CC   (e.g. corneal infections). Other uses include wound healing, maintenance
CC   of organs before transplantation, support of cell culture of primary

  Query Match             25.0%;  Score 278.5;  DB 4;  Length 210;
  Best Local Similarity   37.3%;  Pred. No. 8.8e-19;
  Matches   59;  Conservative   25;  Mismatches   65;  Indels    9;  Gaps    2;

Qy         51 LEVRTSTASGLLLWQGVEVGEAGQGKDFISLGLQDGHLVFRYQLGSGEARLVSEDPINDG 110
              :  :|:   |||||:|          |||||||:|| ||| | |||| | ::     |||
Db         59 MRFKTTAKDGLLLWRG--DSPMRPNSDFISLGLRDGALVFSYNLGSGVASIMVNGSFNDG 116

Qy        111 EWHRVTALREGRRGSIQVDGEELVSGRSPGPNVAVNAKGSVYIGGAPDVATLTGGRFSSG 170
               |||| |:|:|: | | ||     :|:|||    :|  |::|:||  ::|  |  ::  |
Db        117 RWHRVKAVRDGQSGKITVDDYGARTGKSPGMMRQLNINGALYVGGMKEIALHTNRQYMRG 176

Qy        171 ITGCVKNLVLHSARPGAPPPQPLDLQHRAQAGANTRPC 208
              : ||: :  | :          : |   |  | |   |
Db        177 LVGCISHFTLST-------DYHISLVEDAVDGKNINTC 207


RESULT 9
ABP48002
ID   ABP48002 standard; protein; 210 AA.
XX
AC   ABP48002;
XX
DT   23-AUG-2002  (first entry)
XX
DE   Human polypeptide SEQ ID NO 432.
XX
KW   Human; nootropic; neuroprotective; cytostatic; dermatological; virucide;
KW   immunosuppressive; antiinflammatory; anti-HIV; antibacterial; vulnerary;
KW   antiparkinsonian; antisickling; antianaemic; antiarthritic; cancer;
KW   antirheumatic; hepatotropic; cerebroprotective; antiinflammatory;
KW   antiallergic; antidiabetic; antiulcer; anticonvulsant; antifungal;
KW   antiparasitic; cardiant; immune disorder; cardiovascular disorder;
KW   neurological disease; infection; nephrotropic; gene therapy; vaccine.
XX
OS   Homo sapiens.
XX
PN   US2002042386-A1.
XX
PD   11-APR-2002.
XX
PF   17-JAN-2001; 2001US-00764870.
XX
PR   31-JAN-2000; 2000US-0179065P.
PR   04-FEB-2000; 2000US-0180628P.
PR   28-JUN-2000; 2000US-0214886P.
PR   07-JUL-2000; 2000US-0216647P.
PR   07-JUL-2000; 2000US-0216880P.
PR   11-JUL-2000; 2000US-0217487P.
PR   11-JUL-2000; 2000US-0217496P.
PR   14-JUL-2000; 2000US-0218290P.
PR   26-JUL-2000; 2000US-0220963P.
PR   26-JUL-2000; 2000US-0220964P.
```

```
PR   14-AUG-2000; 2000US-0224518P.
PR   14-AUG-2000; 2000US-0224519P.
PR   14-AUG-2000; 2000US-0225267P.
PR   14-AUG-2000; 2000US-0225268P.
PR   14-AUG-2000; 2000US-0225270P.
PR   14-AUG-2000; 2000US-0225447P.
PR   14-AUG-2000; 2000US-0225757P.
PR   14-AUG-2000; 2000US-0225758P.
PR   22-AUG-2000; 2000US-0226868P.
PR   30-AUG-2000; 2000US-0228924P.
PR   01-SEP-2000; 2000US-0229287P.
PR   01-SEP-2000; 2000US-0229343P.
PR   01-SEP-2000; 2000US-0229344P.
PR   01-SEP-2000; 2000US-0229345P.
PR   05-SEP-2000; 2000US-0229509P.
PR   05-SEP-2000; 2000US-0229513P.
PR   08-SEP-2000; 2000US-0231413P.
PR   21-SEP-2000; 2000US-0234223P.
PR   21-SEP-2000; 2000US-0234274P.
PR   25-SEP-2000; 2000US-0234997P.
PR   27-SEP-2000; 2000US-0235834P.
PR   29-SEP-2000; 2000US-0236327P.
PR   29-SEP-2000; 2000US-0236367P.
PR   29-SEP-2000; 2000US-0236368P.
PR   29-SEP-2000; 2000US-0236369P.
PR   29-SEP-2000; 2000US-0236370P.
PR   02-OCT-2000; 2000US-0236802P.
PR   02-OCT-2000; 2000US-0237037P.
PR   02-OCT-2000; 2000US-0237038P.
PR   02-OCT-2000; 2000US-0237039P.
PR   02-OCT-2000; 2000US-0237040P.
PR   13-OCT-2000; 2000US-0239935P.
PR   20-OCT-2000; 2000US-0240960P.
PR   20-OCT-2000; 2000US-0241785P.
PR   20-OCT-2000; 2000US-0241809P.
PR   01-NOV-2000; 2000US-0244617P.
PR   17-NOV-2000; 2000US-0249299P.
PR   08-DEC-2000; 2000US-0251856P.
PR   08-DEC-2000; 2000US-0251868P.
PR   08-DEC-2000; 2000US-0251869P.
XX
PA   (ROSE/) ROSEN C A.
PA   (RUBE/) RUBEN S M.
PA   (BARA/) BARASH S C.
XX
PI   Rosen CA,  Ruben SM,  Barash SC;
XX
DR   WPI; 2002-470713/50.
DR   N-PSDB; ABQ66677.
XX
PT   New nucleic acid encoding human proteins, useful for diagnosis, treatment
PT   and prevention of e.g. osteoporosis, also related polypeptides and
PT   antibodies.
XX
PS   Claim 11; SEQ ID NO 432; 235pp + Sequence Listing; English.
XX
CC   The invention relates to novel genes (ABQ66521-ABQ66785) and proteins
CC   (ABP47846-ABP48110) useful for preventing, treating or ameliorating
CC   medical conditions e.g. by protein or gene therapy. The genes are
CC   isolated from a range of human tissues disclosed in the specification.
CC   The nucleic acids, proteins, antibodies and (ant)agonists are useful in
CC   the diagnosis, treatment and prevention of: (a) cancer, e.g. breast and
CC   ovarian cancer and other cancers of the adrenal gland, bone, bone marrow,
CC   breast, gastrointestinal tract, liver, lung, or urogenital; (b) immune
CC   disorders e.g. Addison's disease, allergies, autoimmune haemolytic
CC   anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's disease,
CC   multiple sclerosis, rheumatoid arthritis and ulcerative colitis; (c)
CC   cardiovascular disorders such as myocardial ischaemias; (d) wound healing
CC   ; (e) neurological diseases e.g. cerebral anoxia and epilepsy; and (f)
CC   infectious diseases such as viral, bacterial, fungal and parasitic
CC   infections. Note: The sequence data for this patent did not form part of
CC   the printed specification, but was obtained in electronic format directly
CC   from USPTO at seqdata.uspto.gov/sequence.html?DocID=999909764870
```

```
XX
SQ   Sequence 210 AA;

  Query Match             25.0%;  Score 278.5;  DB 5;  Length 210;
  Best Local Similarity   37.3%;  Pred. No. 8.8e-19;
  Matches   59;  Conservative   25;  Mismatches   65;  Indels    9;  Gaps    2;

Qy         51 LEVRTSTASGLLLWQGVEVGEAGQGKDFISLGLQDGHLVFRYQLGSGEARLVSEDPINDG 110
              :  :|:   |||||:|          |||||||:|| ||| | |||| | ::     |||
Db         59 MRFKTTAKDGLLLWRG--DSPMRPNSDFISLGLRDGALVFSYNLGSGVASIMVNGSFNDG 116

Qy        111 EWHRVTALREGRRGSIQVDGEELVSGRSPGPNVAVNAKGSVYIGGAPDVATLTGGRFSSG 170
               |||| |:|:|: | | ||     :|:|||    :|  |::|:||  ::|  |  ::  |
Db        117 RWHRVKAVRDGQSGKITVDDYGARTGKSPGMMRQLNINGALYVGGMKEIALHTNRQYMRG 176

Qy        171 ITGCVKNLVLHSARPGAPPPQPLDLQHRAQAGANTRPC 208
              : ||: :  | :          : |   |  | |   |
Db        177 LVGCISHFTLST-------DYHISLVEDAVDGKNINTC 207


RESULT 10
ADC10964
ID   ADC10964 standard; protein; 210 AA.
XX
AC   ADC10964;
XX
DT   18-DEC-2003  (first entry)
XX
DE   Human extracellular matrix protein from gene 157.
XX
KW   Extracellular matrix protein; cytostatic; antibacterial; virucide;
KW   neuroprotective; gynaecological; gastrointestinal-Gen; cardiant;
KW   cardiovascular-Gen; nephrotropic; antiinflammatory; muscular-Gen;
KW   respiratory-Gen; immunosuppressive; cerebroprotective; vasotropic;
KW   nootropic; antiallergic; cancer; bacterial infection; viral infection;
KW   neural disorder; immune system disorder; blood disorder;
KW   muscular disorder; reproductive disorder; gastrointestinal disorder;
KW   pulmonary disorder; cardiovascular disorder; renal disorder;
KW   inflammatory disorder; proliferative disorder; Human.
XX
OS   Homo sapiens.
XX
PN   US2003059875-A1.
XX
PD   27-MAR-2003.
XX
PF   19-APR-2002; 2002US-00125540.
XX
PR   31-JAN-2000; 2000US-0179065P.
PR   04-FEB-2000; 2000US-0180628P.
PR   24-FEB-2000; 2000US-0184664P.
PR   02-MAR-2000; 2000US-0186350P.
PR   16-MAR-2000; 2000US-0189874P.
PR   17-MAR-2000; 2000US-0190076P.
PR   18-APR-2000; 2000US-0198123P.
PR   19-MAY-2000; 2000US-0205515P.
PR   07-JUN-2000; 2000US-0209467P.
PR   28-JUN-2000; 2000US-0214886P.
PR   30-JUN-2000; 2000US-0215135P.
PR   07-JUL-2000; 2000US-0216647P.
PR   07-JUL-2000; 2000US-0216880P.
PR   11-JUL-2000; 2000US-0217487P.
PR   11-JUL-2000; 2000US-0217496P.
PR   14-JUL-2000; 2000US-0218290P.
PR   26-JUL-2000; 2000US-0220963P.
PR   26-JUL-2000; 2000US-0220964P.
PR   14-AUG-2000; 2000US-0224518P.
PR   14-AUG-2000; 2000US-0224519P.
PR   14-AUG-2000; 2000US-0225213P.
PR   14-AUG-2000; 2000US-0225214P.
PR   14-AUG-2000; 2000US-0225266P.
PR   14-AUG-2000; 2000US-0225267P.
PR   14-AUG-2000; 2000US-0225268P.
```

```
PR   14-AUG-2000; 2000US-0225270P.
PR   14-AUG-2000; 2000US-0225447P.
PR   14-AUG-2000; 2000US-0225757P.
PR   14-AUG-2000; 2000US-0225758P.
PR   14-AUG-2000; 2000US-0225759P.
PR   18-AUG-2000; 2000US-0226279P.
PR   22-AUG-2000; 2000US-0226681P.
PR   22-AUG-2000; 2000US-0226868P.
PR   22-AUG-2000; 2000US-0227182P.
PR   23-AUG-2000; 2000US-0227009P.
PR   30-AUG-2000; 2000US-0228924P.
PR   01-SEP-2000; 2000US-0229287P.
PR   01-SEP-2000; 2000US-0229343P.
PR   01-SEP-2000; 2000US-0229344P.
PR   01-SEP-2000; 2000US-0229345P.
PR   05-SEP-2000; 2000US-0229509P.
PR   05-SEP-2000; 2000US-0229513P.
PR   06-SEP-2000; 2000US-0230437P.
PR   06-SEP-2000; 2000US-0230438P.
PR   08-SEP-2000; 2000US-0231242P.
PR   08-SEP-2000; 2000US-0231243P.
PR   08-SEP-2000; 2000US-0231244P.
PR   08-SEP-2000; 2000US-0231413P.
PR   08-SEP-2000; 2000US-0231414P.
PR   08-SEP-2000; 2000US-0232080P.
PR   08-SEP-2000; 2000US-0232081P.
PR   12-SEP-2000; 2000US-0231968P.
PR   14-SEP-2000; 2000US-0232397P.
PR   14-SEP-2000; 2000US-0232398P.
PR   14-SEP-2000; 2000US-0232399P.
PR   14-SEP-2000; 2000US-0232400P.
PR   14-SEP-2000; 2000US-0232401P.
PR   14-SEP-2000; 2000US-0233063P.
PR   14-SEP-2000; 2000US-0233064P.
PR   14-SEP-2000; 2000US-0233065P.
PR   21-SEP-2000; 2000US-0234223P.
PR   21-SEP-2000; 2000US-0234274P.
PR   25-SEP-2000; 2000US-0234997P.
PR   25-SEP-2000; 2000US-0234998P.
PR   26-SEP-2000; 2000US-0235484P.
PR   27-SEP-2000; 2000US-0235834P.
PR   27-SEP-2000; 2000US-0235836P.
PR   29-SEP-2000; 2000US-0236327P.
PR   29-SEP-2000; 2000US-0236367P.
PR   29-SEP-2000; 2000US-0236368P.
PR   29-SEP-2000; 2000US-0236369P.
PR   29-SEP-2000; 2000US-0236370P.
PR   02-OCT-2000; 2000US-0236802P.
PR   02-OCT-2000; 2000US-0237037P.
PR   02-OCT-2000; 2000US-0237038P.
PR   02-OCT-2000; 2000US-0237039P.
PR   02-OCT-2000; 2000US-0237040P.
PR   13-OCT-2000; 2000US-0239935P.
PR   13-OCT-2000; 2000US-0239937P.
PR   20-OCT-2000; 2000US-0240960P.
PR   20-OCT-2000; 2000US-0241221P.
PR   20-OCT-2000; 2000US-0241785P.
PR   20-OCT-2000; 2000US-0241786P.
PR   20-OCT-2000; 2000US-0241787P.
PR   20-OCT-2000; 2000US-0241808P.
PR   20-OCT-2000; 2000US-0241809P.
PR   20-OCT-2000; 2000US-0241826P.
PR   01-NOV-2000; 2000US-0244617P.
PR   08-NOV-2000; 2000US-0246474P.
PR   08-NOV-2000; 2000US-0246475P.
PR   08-NOV-2000; 2000US-0246476P.
PR   08-NOV-2000; 2000US-0246477P.
PR   08-NOV-2000; 2000US-0246478P.
PR   08-NOV-2000; 2000US-0246523P.
PR   08-NOV-2000; 2000US-0246524P.
PR   08-NOV-2000; 2000US-0246525P.
PR   08-NOV-2000; 2000US-0246526P.
PR   08-NOV-2000; 2000US-0246527P.
```

```
PR   08-NOV-2000; 2000US-0246528P.
PR   08-NOV-2000; 2000US-0246532P.
PR   08-NOV-2000; 2000US-0246609P.
PR   08-NOV-2000; 2000US-0246610P.
PR   08-NOV-2000; 2000US-0246611P.
PR   08-NOV-2000; 2000US-0246613P.
PR   17-NOV-2000; 2000US-0249207P.
PR   17-NOV-2000; 2000US-0249208P.
PR   17-NOV-2000; 2000US-0249209P.
PR   17-NOV-2000; 2000US-0249210P.
PR   17-NOV-2000; 2000US-0249211P.
PR   17-NOV-2000; 2000US-0249212P.
PR   17-NOV-2000; 2000US-0249213P.
PR   17-NOV-2000; 2000US-0249214P.
PR   17-NOV-2000; 2000US-0249215P.
PR   17-NOV-2000; 2000US-0249216P.
PR   17-NOV-2000; 2000US-0249217P.
PR   17-NOV-2000; 2000US-0249218P.
PR   17-NOV-2000; 2000US-0249244P.
PR   17-NOV-2000; 2000US-0249245P.
PR   17-NOV-2000; 2000US-0249264P.
PR   17-NOV-2000; 2000US-0249265P.
PR   17-NOV-2000; 2000US-0249297P.
PR   17-NOV-2000; 2000US-0249299P.
PR   17-NOV-2000; 2000US-0249300P.
PR   01-DEC-2000; 2000US-0250160P.
PR   01-DEC-2000; 2000US-0250391P.
PR   05-DEC-2000; 2000US-0251030P.
PR   05-DEC-2000; 2000US-0251988P.
PR   05-DEC-2000; 2000US-0256719P.
PR   06-DEC-2000; 2000US-0251479P.
PR   08-DEC-2000; 2000US-0251856P.
PR   08-DEC-2000; 2000US-0251868P.
PR   08-DEC-2000; 2000US-0251869P.
PR   08-DEC-2000; 2000US-0251989P.
PR   08-DEC-2000; 2000US-0251990P.
PR   11-DEC-2000; 2000US-0254097P.
PR   05-JAN-2001; 2001US-0259678P.
PR   17-JAN-2001; 2001US-00764870.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Rosen CA,  Ruben SM,  Barash SC;
XX
DR   WPI; 2003-743765/70.
DR   N-PSDB; ADC10699.
XX
PT   New isolated nucleic acids and polypeptides, useful for diagnosing,
PT   treating, and/or preventing disorders, such as cancer, infections,
PT   cardiovascular and inflammatory diseases.
XX
PS   Claim 11; SEQ ID NO 432; 235pp; English.
XX
CC   The invention relates to an isolated nucleic acid molecule (cDNA)
CC   encoding a human extracellular matrix protein, representing one of 161
CC   novel genes. Also included are recombinant vectors, host cells
CC   (expressing the protein), the extracellular matrix proteins (including
CC   their fragments, epitopes and homologues), an isolated antibody that
CC   binds specifically to the protein, diagnosing a pathological condition or
CC   susceptibility to a pathological condition (comprising determining the
CC   presence or absence of a mutation in the nucleic acid and diagnosing a
CC   condition based on the presence or absence of the mutation), diagnosing a
CC   pathological condition or susceptibility to a pathological condition
CC   (comprising determining the presence or amount of expression of the
CC   protein in a biological sample and diagnosing a condition based on the
CC   presence or amount of expression of the protein), preventing, treating or
CC   ameliorating a medical condition by administering the nucleic acid or
CC   protein to a mammalian subject, identifying a binding partner to the
CC   protein, the gene corresponding to the cDNA sequence, and identifying an
CC   activity in a biological assay (comprising expressing the nucleic acid in
CC   a cell, isolating the supernatant, detecting an activity in a biological
CC   assay and identifying the protein in the supernatant having the
CC   activity). The nucleic acids and proteins display the following
```

```
CC   activities Cytostatic, antibacterial, Virucide, Neuroprotective,
CC   Gynaecological, Gastrointestinal-Gen, Cardiant, Cardiovascular-Gen,
CC   Nephrotropic, Antiinflammatory, Muscular-Gen, Respiratory-Gen,

  Query Match              25.0%;  Score 278.5;  DB 7;  Length 210;
  Best Local Similarity   37.3%;  Pred. No. 8.8e-19;
  Matches   59;  Conservative   25;  Mismatches   65;  Indels    9;  Gaps    2;

Qy         51 LEVRTSTASGLLLWQGVEVGEAGQGKDFISLGLQDGHLVFRYQLGSGEARLVSEDPINDG 110
              :  :|:   |||||:|          |||||||:|| ||| | |||| | ::     |||
Db         59 MRFKTTAKDGLLLWRG--DSPMRPNSDFISLGLRDGALVFSYNLGSGVASIMVNGSFNDG 116

Qy        111 EWHRVTALREGRRGSIQVDGEELVSGRSPGPNVAVNAKGSVYIGGAPDVATLTGGRFSSG 170
               |||| |:|:|: | | ||     :|:|||    :|  |::|:||  ::|  |   :: |
Db        117 RWHRVKAVRDGQSGKITVDDYGARTGKSPGMMRQLNINGALYVGGMKEIALHTNRQYMRG 176

Qy        171 ITGCVKNLVLHSARPGAPPPQPLDLQHRAQAGANTRPC 208
              : ||: :  | :          : |   |  | |   |
Db        177 LVGCISHFTLST-------DYHISLVEDAVDGKNINTC 207


RESULT 11
AAU18145
ID   AAU18145 standard; protein; 238 AA.
XX
AC   AAU18145;
XX
DT   21-NOV-2001  (first entry)
XX
DE   Novel human uterine motility-association polypeptide #52.
XX
KW   Human; uterine motility-association disorder; uterus; pregnancy; labour;
KW   menstrual cycle; gene therapy.
XX
OS   Homo sapiens.
XX
PN   WO200155201-A1.
XX
PD   02-AUG-2001.
XX
PF   17-JAN-2001; 2001WO-US001317.
XX
PR   31-JAN-2000; 2000US-0179065P.
PR   04-FEB-2000; 2000US-0180628P.
PR   24-FEB-2000; 2000US-0184664P.
PR   02-MAR-2000; 2000US-0186350P.
PR   16-MAR-2000; 2000US-0189874P.
PR   17-MAR-2000; 2000US-0190076P.
PR   18-APR-2000; 2000US-0198123P.
PR   19-MAY-2000; 2000US-0205515P.
PR   07-JUN-2000; 2000US-0209467P.
PR   28-JUN-2000; 2000US-0214886P.
PR   30-JUN-2000; 2000US-0215135P.
PR   07-JUL-2000; 2000US-0216647P.
PR   07-JUL-2000; 2000US-0216880P.
PR   11-JUL-2000; 2000US-0217487P.
PR   11-JUL-2000; 2000US-0217496P.
PR   14-JUL-2000; 2000US-0218290P.
PR   26-JUL-2000; 2000US-0220963P.
PR   26-JUL-2000; 2000US-0220964P.
PR   14-AUG-2000; 2000US-0224518P.
PR   14-AUG-2000; 2000US-0224519P.
PR   14-AUG-2000; 2000US-0225213P.
PR   14-AUG-2000; 2000US-0225214P.
PR   14-AUG-2000; 2000US-0225266P.
PR   14-AUG-2000; 2000US-0225267P.
PR   14-AUG-2000; 2000US-0225268P.
PR   14-AUG-2000; 2000US-0225270P.
PR   14-AUG-2000; 2000US-0225447P.
PR   14-AUG-2000; 2000US-0225757P.
PR   14-AUG-2000; 2000US-0225758P.
PR   14-AUG-2000; 2000US-0225759P.
PR   18-AUG-2000; 2000US-0226279P.
PR   22-AUG-2000; 2000US-0226681P.
PR   22-AUG-2000; 2000US-0226868P.
PR   22-AUG-2000; 2000US-0227182P.
PR   23-AUG-2000; 2000US-0227009P.
PR   30-AUG-2000; 2000US-0228924P.
PR   01-SEP-2000; 2000US-0229287P.
PR   01-SEP-2000; 2000US-0229343P.
PR   01-SEP-2000; 2000US-0229344P.
PR   01-SEP-2000; 2000US-0229345P.
PR   05-SEP-2000; 2000US-0229509P.
PR   05-SEP-2000; 2000US-0229513P.
PR   06-SEP-2000; 2000US-0230437P.
PR   06-SEP-2000; 2000US-0230438P.
PR   08-SEP-2000; 2000US-0231242P.
PR   08-SEP-2000; 2000US-0231243P.
PR   08-SEP-2000; 2000US-0231244P.
PR   08-SEP-2000; 2000US-0231413P.
PR   08-SEP-2000; 2000US-0231414P.
PR   08-SEP-2000; 2000US-0232080P.
PR   08-SEP-2000; 2000US-0232081P.
PR   12-SEP-2000; 2000US-0231968P.
PR   14-SEP-2000; 2000US-0232397P.
PR   14-SEP-2000; 2000US-0232398P.
PR   14-SEP-2000; 2000US-0232399P.
PR   14-SEP-2000; 2000US-0232400P.
PR   14-SEP-2000; 2000US-0232401P.
PR   14-SEP-2000; 2000US-0233063P.
PR   14-SEP-2000; 2000US-0233064P.
PR   14-SEP-2000; 2000US-0233065P.
PR   21-SEP-2000; 2000US-0234223P.
PR   21-SEP-2000; 2000US-0234274P.
PR   25-SEP-2000; 2000US-0234997P.
PR   25-SEP-2000; 2000US-0234998P.
PR   26-SEP-2000; 2000US-0235484P.
PR   27-SEP-2000; 2000US-0235834P.
PR   27-SEP-2000; 2000US-0235836P.
PR   29-SEP-2000; 2000US-0236327P.
PR   29-SEP-2000; 2000US-0236367P.
PR   29-SEP-2000; 2000US-0236368P.
PR   29-SEP-2000; 2000US-0236369P.
PR   29-SEP-2000; 2000US-0236370P.
PR   02-OCT-2000; 2000US-0236802P.
PR   02-OCT-2000; 2000US-0237037P.
PR   02-OCT-2000; 2000US-0237038P.
PR   02-OCT-2000; 2000US-0237039P.
PR   02-OCT-2000; 2000US-0237040P.
PR   13-OCT-2000; 2000US-0239935P.
PR   13-OCT-2000; 2000US-0239937P.
PR   20-OCT-2000; 2000US-0240960P.
PR   20-OCT-2000; 2000US-0241221P.
PR   20-OCT-2000; 2000US-0241785P.
PR   20-OCT-2000; 2000US-0241786P.
PR   20-OCT-2000; 2000US-0241787P.
PR   20-OCT-2000; 2000US-0241808P.
PR   20-OCT-2000; 2000US-0241809P.
PR   20-OCT-2000; 2000US-0241826P.
PR   01-NOV-2000; 2000US-0244617P.
PR   08-NOV-2000; 2000US-0246474P.
PR   08-NOV-2000; 2000US-0246475P.
PR   08-NOV-2000; 2000US-0246476P.
PR   08-NOV-2000; 2000US-0246477P.
PR   08-NOV-2000; 2000US-0246478P.
PR   08-NOV-2000; 2000US-0246523P.
PR   08-NOV-2000; 2000US-0246524P.
PR   08-NOV-2000; 2000US-0246525P.
PR   08-NOV-2000; 2000US-0246526P.
PR   08-NOV-2000; 2000US-0246527P.
PR   08-NOV-2000; 2000US-0246528P.
PR   08-NOV-2000; 2000US-0246532P.
PR   08-NOV-2000; 2000US-0246609P.
PR   08-NOV-2000; 2000US-0246610P.
PR   08-NOV-2000; 2000US-0246611P.
PR   08-NOV-2000; 2000US-0246613P.
```

```
PR   17-NOV-2000; 2000US-0249207P.
PR   17-NOV-2000; 2000US-0249208P.
PR   17-NOV-2000; 2000US-0249209P.
PR   17-NOV-2000; 2000US-0249210P.
PR   17-NOV-2000; 2000US-0249211P.
PR   17-NOV-2000; 2000US-0249212P.
PR   17-NOV-2000; 2000US-0249213P.
PR   17-NOV-2000; 2000US-0249214P.
PR   17-NOV-2000; 2000US-0249215P.
PR   17-NOV-2000; 2000US-0249216P.
PR   17-NOV-2000; 2000US-0249217P.
PR   17-NOV-2000; 2000US-0249218P.
PR   17-NOV-2000; 2000US-0249244P.
PR   17-NOV-2000; 2000US-0249245P.
PR   17-NOV-2000; 2000US-0249264P.
PR   17-NOV-2000; 2000US-0249265P.
PR   17-NOV-2000; 2000US-0249297P.
PR   17-NOV-2000; 2000US-0249299P.
PR   17-NOV-2000; 2000US-0249300P.
PR   01-DEC-2000; 2000US-0250160P.
PR   01-DEC-2000; 2000US-0250391P.
PR   05-DEC-2000; 2000US-0251030P.
PR   05-DEC-2000; 2000US-0251988P.
PR   05-DEC-2000; 2000US-0256719P.
PR   06-DEC-2000; 2000US-0251479P.
PR   08-DEC-2000; 2000US-0251856P.
PR   08-DEC-2000; 2000US-0251868P.
PR   08-DEC-2000; 2000US-0251869P.
PR   08-DEC-2000; 2000US-0251989P.
PR   08-DEC-2000; 2000US-0251990P.
PR   11-DEC-2000; 2000US-0254097P.
PR   05-JAN-2001; 2001US-0259678P.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Rosen CA,  Barash SC,  Ruben SM;
XX
DR   WPI; 2001-488777/53.
DR   N-PSDB; AAS28987.
XX
PT   New nucleic acid molecules encoding 49 human secreted proteins for
PT   diagnosing, preventing, treating or ameliorating medical conditions and
PT   used as food additives or preservatives.
XX
PS   Claim 11; SEQ ID NO 121; 524pp; English.
XX
CC   The present invention relates to the isolation of novel human uterine
CC   motility-association polypeptides, and cDNA (AAS28936-AAS28994) and
CC   genomic sequences encoding for these polypeptides. The sequences of the
CC   invention are useful in the diagnosis, treatment, prevention and/or
CC   prognosis of diseases associated with uterine motility such as pregnancy
CC   and labour, and menstrual disorders. The polynucleotide sequences of the
CC   invention are also useful in gene therapy. AAU18094-AAU18152 represent
CC   novel human uterine motility-association polypeptides. Note: The sequence
CC   data for this patent did not form part of the printed specification, but
CC   was obtained in electronic format directly from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 238 AA;

  Query Match             25.0%;  Score 278.5;  DB 4;  Length 238;
  Best Local Similarity   37.3%;  Pred. No. 1e-18;
  Matches   59;  Conservative   25;  Mismatches   65;  Indels    9;  Gaps    2;

Qy         51 LEVRTSTASGLLLWQGVEVGEAGQGKDFISLGLQDGHLVFRYQLGSGEARLVSEDPINDG 110
              :  :|:   |||||:|          |||||||:|| ||| | |||| | ::     |||
Db         87 MRFKTTAKDGLLLWRG--DSPMRPNSDFISLGLRDGALVFSYNLGSGVASIMVNGSFNDG 144

Qy        111 EWHRVTALREGRRGSIQVDGEELVSGRSPGPNVAVNAKGSVYIGGAPDVATLTGGRFSSG 170
               |||| |:|:|: | | ||     :|:|||    :|  |::|:||  ::|  |  ::  |
Db        145 RWHRVKAVRDGQSGKITVDDYGARTGKSPGMMRQLNINGALYVGGMKEIALHTNRQYMRG 204

Qy        171 ITGCVKNLVLHSARPGAPPPQPLDLQHRAQAGANTRPC 208
              : ||: :  | :          : |   |  | |   |
Db        205 LVGCISHFTLST-------DYHISLVEDAVDGKNINTC 235

RESULT 12
AAU17011
ID   AAU17011 standard; protein; 238 AA.
XX
AC   AAU17011;
XX
DT   07-NOV-2001  (first entry)
XX
DE   Human novel secreted protein, SEQ ID 252.
XX
KW   Human; immunosuppressive; antiarthritic; antirheumatic; cytostatic;
KW   cardiant; vasotropic; cerebroprotective; nootropic; neuroprotective;
KW   antibacterial; virucide; fungicide; opthalmalogical; vulnerary;
KW   secreted protein; rheumatoid arthritis; hyperproliferative disorder;
KW   cardiovascular disorder; cardiac arrest; cerebrovascular disorder;
KW   cerebral ischaemia; angiogenesis; nervous system disorder;
KW   Alzheimer's disease; infection; ocular disorder; corneal infection;
KW   wound healing; epithelial cell proliferation; skin ageing; food additive;
KW   preservative; antiproliferative.
XX
OS   Homo sapiens.
XX
PN   WO200155441-A2.
XX
PD   02-AUG-2001.
XX
PF   17-JAN-2001; 2001WO-US001320.
XX
PR   31-JAN-2000; 2000US-0179065P.
PR   04-FEB-2000; 2000US-0180628P.
PR   24-FEB-2000; 2000US-0184664P.
PR   02-MAR-2000; 2000US-0186350P.
PR   16-MAR-2000; 2000US-0189874P.
PR   17-MAR-2000; 2000US-0190076P.
PR   18-APR-2000; 2000US-0198123P.
PR   19-MAY-2000; 2000US-0205515P.
PR   07-JUN-2000; 2000US-0209467P.
PR   28-JUN-2000; 2000US-0214886P.
PR   30-JUN-2000; 2000US-0215135P.
PR   07-JUL-2000; 2000US-0216647P.
PR   07-JUL-2000; 2000US-0216880P.
PR   11-JUL-2000; 2000US-0217487P.
PR   11-JUL-2000; 2000US-0217496P.
PR   14-JUL-2000; 2000US-0218290P.
PR   26-JUL-2000; 2000US-0220963P.
PR   26-JUL-2000; 2000US-0220964P.
PR   14-AUG-2000; 2000US-0224518P.
PR   14-AUG-2000; 2000US-0224519P.
PR   14-AUG-2000; 2000US-0225213P.
PR   14-AUG-2000; 2000US-0225214P.
PR   14-AUG-2000; 2000US-0225266P.
PR   14-AUG-2000; 2000US-0225267P.
PR   14-AUG-2000; 2000US-0225268P.
PR   14-AUG-2000; 2000US-0225270P.
PR   14-AUG-2000; 2000US-0225447P.
PR   14-AUG-2000; 2000US-0225757P.
PR   14-AUG-2000; 2000US-0225758P.
PR   14-AUG-2000; 2000US-0225759P.
PR   18-AUG-2000; 2000US-0226279P.
PR   22-AUG-2000; 2000US-0226681P.
PR   22-AUG-2000; 2000US-0226868P.
PR   22-AUG-2000; 2000US-0227182P.
PR   23-AUG-2000; 2000US-0227009P.
PR   30-AUG-2000; 2000US-0228924P.
PR   01-SEP-2000; 2000US-0229287P.
PR   01-SEP-2000; 2000US-0229343P.
PR   01-SEP-2000; 2000US-0229344P.
PR   01-SEP-2000; 2000US-0229345P.
PR   05-SEP-2000; 2000US-0229509P.
```

```
PR   05-SEP-2000; 2000US-0229513P.
PR   06-SEP-2000; 2000US-0230437P.
PR   06-SEP-2000; 2000US-0230438P.
PR   08-SEP-2000; 2000US-0231242P.
PR   08-SEP-2000; 2000US-0231243P.
PR   08-SEP-2000; 2000US-0231244P.
PR   08-SEP-2000; 2000US-0231413P.
PR   08-SEP-2000; 2000US-0231414P.
PR   08-SEP-2000; 2000US-0232080P.
PR   08-SEP-2000; 2000US-0232081P.
PR   12-SEP-2000; 2000US-0231968P.
PR   14-SEP-2000; 2000US-0232397P.
PR   14-SEP-2000; 2000US-0232398P.
PR   14-SEP-2000; 2000US-0232399P.
PR   14-SEP-2000; 2000US-0232400P.
PR   14-SEP-2000; 2000US-0232401P.
PR   14-SEP-2000; 2000US-0233063P.
PR   14-SEP-2000; 2000US-0233064P.
PR   14-SEP-2000; 2000US-0233065P.
PR   21-SEP-2000; 2000US-0234223P.
PR   21-SEP-2000; 2000US-0234274P.
PR   25-SEP-2000; 2000US-0234997P.
PR   25-SEP-2000; 2000US-0234998P.
PR   26-SEP-2000; 2000US-0235484P.
PR   27-SEP-2000; 2000US-0235834P.
PR   27-SEP-2000; 2000US-0235836P.
PR   29-SEP-2000; 2000US-0236327P.
PR   29-SEP-2000; 2000US-0236367P.
PR   29-SEP-2000; 2000US-0236368P.
PR   29-SEP-2000; 2000US-0236369P.
PR   29-SEP-2000; 2000US-0236370P.
PR   02-OCT-2000; 2000US-0236802P.
PR   02-OCT-2000; 2000US-0237037P.
PR   02-OCT-2000; 2000US-0237038P.
PR   02-OCT-2000; 2000US-0237039P.
PR   02-OCT-2000; 2000US-0237040P.
PR   13-OCT-2000; 2000US-0239935P.
PR   13-OCT-2000; 2000US-0239937P.
PR   20-OCT-2000; 2000US-0240960P.
PR   20-OCT-2000; 2000US-0241221P.
PR   20-OCT-2000; 2000US-0241785P.
PR   20-OCT-2000; 2000US-0241786P.
PR   20-OCT-2000; 2000US-0241787P.
PR   20-OCT-2000; 2000US-0241808P.
PR   20-OCT-2000; 2000US-0241809P.
PR   20-OCT-2000; 2000US-0241826P.
PR   01-NOV-2000; 2000US-0244617P.
PR   08-NOV-2000; 2000US-0246474P.
PR   08-NOV-2000; 2000US-0246475P.
PR   08-NOV-2000; 2000US-0246476P.
PR   08-NOV-2000; 2000US-0246477P.
PR   08-NOV-2000; 2000US-0246478P.
PR   08-NOV-2000; 2000US-0246523P.
PR   08-NOV-2000; 2000US-0246524P.
PR   08-NOV-2000; 2000US-0246525P.
PR   08-NOV-2000; 2000US-0246526P.
PR   08-NOV-2000; 2000US-0246527P.
PR   08-NOV-2000; 2000US-0246528P.
PR   08-NOV-2000; 2000US-0246532P.
PR   08-NOV-2000; 2000US-0246609P.
PR   08-NOV-2000; 2000US-0246610P.
PR   08-NOV-2000; 2000US-0246611P.
PR   08-NOV-2000; 2000US-0246613P.
PR   17-NOV-2000; 2000US-0249207P.
PR   17-NOV-2000; 2000US-0249208P.
PR   17-NOV-2000; 2000US-0249209P.
PR   17-NOV-2000; 2000US-0249210P.
PR   17-NOV-2000; 2000US-0249211P.
PR   17-NOV-2000; 2000US-0249212P.
PR   17-NOV-2000; 2000US-0249213P.
PR   17-NOV-2000; 2000US-0249214P.
PR   17-NOV-2000; 2000US-0249215P.
PR   17-NOV-2000; 2000US-0249216P.
```

```
PR   17-NOV-2000; 2000US-0249217P.
PR   17-NOV-2000; 2000US-0249218P.
PR   17-NOV-2000; 2000US-0249244P.
PR   17-NOV-2000; 2000US-0249245P.
PR   17-NOV-2000; 2000US-0249264P.
PR   17-NOV-2000; 2000US-0249265P.
PR   17-NOV-2000; 2000US-0249297P.
PR   17-NOV-2000; 2000US-0249299P.
PR   17-NOV-2000; 2000US-0249300P.
PR   01-DEC-2000; 2000US-0250160P.
PR   01-DEC-2000; 2000US-0250391P.
PR   05-DEC-2000; 2000US-0251030P.
PR   05-DEC-2000; 2000US-0251988P.
PR   05-DEC-2000; 2000US-0256719P.
PR   06-DEC-2000; 2000US-0251479P.
PR   08-DEC-2000; 2000US-0251856P.
PR   08-DEC-2000; 2000US-0251868P.
PR   08-DEC-2000; 2000US-0251869P.
PR   08-DEC-2000; 2000US-0251989P.
PR   08-DEC-2000; 2000US-0251990P.
PR   11-DEC-2000; 2000US-0254097P.
PR   05-JAN-2001; 2001US-0259678P.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Rosen CA,  Barash SC,  Ruben SM;
XX
DR   WPI; 2001-476222/51.
DR   N-PSDB; AAS26916.
XX
PT   Novel polypeptides and polynucleotides useful as diagnostic reagents to
PT   diagnose diseases or disorders associated with aberrant expression or
PT   activity of polypeptides, for treating blood clotting disorder,
PT   hemophilia.
XX
PS   Claim 11; SEQ ID NO 252; 601pp; English.
XX
CC   The invention relates to isolated nucleic acid molecules and their
CC   encoded secreted proteins. The nucleic acids and proteins are used to
CC   prevent, treat or ameliorate a medical condition in e.g. humans, mice,
CC   rabbits, goats, horses, cats, dogs, chickens or sheep. They are also used
CC   in diagnosing a pathological condition or susceptibility to a
CC   pathological condition. Antibodies to the proteins can also be used in
CC   alleviating symptoms associated with the disorders and in diagnostic
CC   immunoassays e.g. radioimmunoassays or enzyme linked immunosorbant assays
CC   (ELISA). Disorders which are diagnosed or treated include autoimmune
CC   diseases e.g. rheumatoid arthritis, hyperproliferative disorders e.g.
CC   neoplasms of the breast or liver, cardiovascular disorders e.g. cardiac
CC   arrest, cerebrovascular disorders e.g. cerebral ischaemia, angiogenesis,
CC   nervous system disorders e.g. Alzheimer's disease, infections caused by
CC   bacteria, viruses and fungi and ocular disorders e.g. corneal infection,
CC   and many other disorders listed in the specification. The polypeptides
CC   can also be used to aid wound healing and epithelial cell proliferation,
CC   to prevent skin aging due to sunburn, to maintain organs before
CC   transplantation, for supporting cell culture of primary tissues, to
CC   regenerate tissues and in chemotaxis. The polypeptides can also be used
CC   as a food additive or preservative to increase or decrease storage
CC   capabilities, fat content, lipid, protein, carbohydrate, vitamins,
CC   minerals, cofactors and other nutritional components. The present
CC   sequence represents a novel secreted protein of the invention. Note: The

  Query Match             25.0%;  Score 278.5;  DB 4;  Length 238;
  Best Local Similarity   37.3%;  Pred. No. 1e-18;
  Matches   59;  Conservative   25;  Mismatches   65;  Indels    9;  Gaps    2;

Qy         51 LEVRTSTASGLLLWQGVEVGEAGQGKDFISLGLQDGHLVFRYQLGSGEARLVSEDPINDG 110
              :  :|:   |||||:|          |||||||:|| ||| | |||| | ::     |||
Db         87 MRFKTTAKDGLLLWRG--DSPMRPNSDFISLGLRDGALVFSYNLGSGVASIMVNGSFNDG 144

Qy        111 EWHRVTALREGRRGSIQVDGEELVSGRSPGPNVAVNAKGSVYIGGAPDVATLTGGRFSSG 170
               |||| |:|:|: | | ||     :|:|||    :|  |::|:||  ::|  |  ::  |
Db        145 RWHRVKAVRDGQSGKITVDDYGARTGKSPGMMRQLNINGALYVGGMKEIALHTNRQYMRG 204
```

```
Qy          171 ITGCVKNLVLHSARPGAPPPQPLDLQHRAQAGANTRPC 208
                : ||: :  | :          : |   |  | |   |
Db          205 LVGCISHFTLST-------DYHISLVEDAVDGKNINTC 235

RESULT 13
ABB10450
ID   ABB10450 standard; protein; 238 AA.
XX
AC   ABB10450;
XX
DT   10-JAN-2002  (first entry)
XX
DE   Human cDNA SEQ ID NO: 758.
XX
KW   Human; gene therapy; neural disorder; immune system disorder;
KW   muscular disorder; reproductive disorder; gastrointestinal disorder;
KW   pulmonary disorder; cardiovascular disorder; renal disorder;
KW   proliferative disorder; inflammation.
XX
OS   Homo sapiens.
XX
PN   WO200154474-A2.
XX
PD   02-AUG-2001.
XX
PF   17-JAN-2001; 2001WO-US001349.
XX
PR   31-JAN-2000; 2000US-0179065P.
PR   04-FEB-2000; 2000US-0180628P.
PR   24-FEB-2000; 2000US-0184664P.
PR   02-MAR-2000; 2000US-0186350P.
PR   16-MAR-2000; 2000US-0189874P.
PR   17-MAR-2000; 2000US-0190076P.
PR   18-APR-2000; 2000US-0198123P.
PR   19-MAY-2000; 2000US-0205515P.
PR   07-JUN-2000; 2000US-0209467P.
PR   28-JUN-2000; 2000US-0214886P.
PR   30-JUN-2000; 2000US-0215135P.
PR   07-JUL-2000; 2000US-0216647P.
PR   07-JUL-2000; 2000US-0216880P.
PR   11-JUL-2000; 2000US-0217487P.
PR   11-JUL-2000; 2000US-0217496P.
PR   14-JUL-2000; 2000US-0218290P.
PR   26-JUL-2000; 2000US-0220963P.
PR   26-JUL-2000; 2000US-0220964P.
PR   14-AUG-2000; 2000US-0224518P.
PR   14-AUG-2000; 2000US-0224519P.
PR   14-AUG-2000; 2000US-0225213P.
PR   14-AUG-2000; 2000US-0225214P.
PR   14-AUG-2000; 2000US-0225266P.
PR   14-AUG-2000; 2000US-0225267P.
PR   14-AUG-2000; 2000US-0225268P.
PR   14-AUG-2000; 2000US-0225270P.
PR   14-AUG-2000; 2000US-0225447P.
PR   14-AUG-2000; 2000US-0225757P.
PR   14-AUG-2000; 2000US-0225758P.
PR   14-AUG-2000; 2000US-0225759P.
PR   18-AUG-2000; 2000US-0226279P.
PR   22-AUG-2000; 2000US-0226681P.
PR   22-AUG-2000; 2000US-0226868P.
PR   22-AUG-2000; 2000US-0227182P.
PR   23-AUG-2000; 2000US-0227009P.
PR   30-AUG-2000; 2000US-0228924P.
PR   01-SEP-2000; 2000US-0229287P.
PR   01-SEP-2000; 2000US-0229343P.
PR   01-SEP-2000; 2000US-0229344P.
PR   01-SEP-2000; 2000US-0229345P.
PR   05-SEP-2000; 2000US-0229509P.
PR   05-SEP-2000; 2000US-0229513P.
PR   06-SEP-2000; 2000US-0230437P.
PR   06-SEP-2000; 2000US-0230438P.
PR   08-SEP-2000; 2000US-0231242P.
```

```
PR   08-SEP-2000; 2000US-0231243P.
PR   08-SEP-2000; 2000US-0231244P.
PR   08-SEP-2000; 2000US-0231413P.
PR   08-SEP-2000; 2000US-0231414P.
PR   08-SEP-2000; 2000US-0232080P.
PR   08-SEP-2000; 2000US-0232081P.
PR   12-SEP-2000; 2000US-0231968P.
PR   14-SEP-2000; 2000US-0232397P.
PR   14-SEP-2000; 2000US-0232398P.
PR   14-SEP-2000; 2000US-0232399P.
PR   14-SEP-2000; 2000US-0232400P.
PR   14-SEP-2000; 2000US-0232401P.
PR   14-SEP-2000; 2000US-0233063P.
PR   14-SEP-2000; 2000US-0233064P.
PR   14-SEP-2000; 2000US-0233065P.
PR   21-SEP-2000; 2000US-0234223P.
PR   21-SEP-2000; 2000US-0234274P.
PR   25-SEP-2000; 2000US-0234997P.
PR   25-SEP-2000; 2000US-0234998P.
PR   26-SEP-2000; 2000US-0235484P.
PR   27-SEP-2000; 2000US-0235834P.
PR   27-SEP-2000; 2000US-0235836P.
PR   29-SEP-2000; 2000US-0236327P.
PR   29-SEP-2000; 2000US-0236367P.
PR   29-SEP-2000; 2000US-0236368P.
PR   29-SEP-2000; 2000US-0236369P.
PR   29-SEP-2000; 2000US-0236370P.
PR   02-OCT-2000; 2000US-0236802P.
PR   02-OCT-2000; 2000US-0237037P.
PR   02-OCT-2000; 2000US-0237038P.
PR   02-OCT-2000; 2000US-0237039P.
PR   02-OCT-2000; 2000US-0237040P.
PR   13-OCT-2000; 2000US-0239935P.
PR   13-OCT-2000; 2000US-0239937P.
PR   20-OCT-2000; 2000US-0240960P.
PR   20-OCT-2000; 2000US-0241221P.
PR   20-OCT-2000; 2000US-0241785P.
PR   20-OCT-2000; 2000US-0241786P.
PR   20-OCT-2000; 2000US-0241787P.
PR   20-OCT-2000; 2000US-0241808P.
PR   20-OCT-2000; 2000US-0241809P.
PR   20-OCT-2000; 2000US-0241826P.
PR   01-NOV-2000; 2000US-0244617P.
PR   08-NOV-2000; 2000US-0246474P.
PR   08-NOV-2000; 2000US-0246475P.
PR   08-NOV-2000; 2000US-0246476P.
PR   08-NOV-2000; 2000US-0246477P.
PR   08-NOV-2000; 2000US-0246478P.
PR   08-NOV-2000; 2000US-0246523P.
PR   08-NOV-2000; 2000US-0246524P.
PR   08-NOV-2000; 2000US-0246525P.
PR   08-NOV-2000; 2000US-0246526P.
PR   08-NOV-2000; 2000US-0246527P.
PR   08-NOV-2000; 2000US-0246528P.
PR   08-NOV-2000; 2000US-0246532P.
PR   08-NOV-2000; 2000US-0246609P.
PR   08-NOV-2000; 2000US-0246610P.
PR   08-NOV-2000; 2000US-0246611P.
PR   08-NOV-2000; 2000US-0246613P.
PR   17-NOV-2000; 2000US-0249207P.
PR   17-NOV-2000; 2000US-0249208P.
PR   17-NOV-2000; 2000US-0249209P.
PR   17-NOV-2000; 2000US-0249210P.
PR   17-NOV-2000; 2000US-0249211P.
PR   17-NOV-2000; 2000US-0249212P.
PR   17-NOV-2000; 2000US-0249213P.
PR   17-NOV-2000; 2000US-0249214P.
PR   17-NOV-2000; 2000US-0249215P.
PR   17-NOV-2000; 2000US-0249216P.
PR   17-NOV-2000; 2000US-0249217P.
PR   17-NOV-2000; 2000US-0249218P.
PR   17-NOV-2000; 2000US-0249244P.
PR   17-NOV-2000; 2000US-0249245P.
```

```
PR   17-NOV-2000; 2000US-0249264P.
PR   17-NOV-2000; 2000US-0249265P.
PR   17-NOV-2000; 2000US-0249297P.
PR   17-NOV-2000; 2000US-0249299P.
PR   17-NOV-2000; 2000US-0249300P.
PR   01-DEC-2000; 2000US-0250160P.
PR   01-DEC-2000; 2000US-0250391P.
PR   05-DEC-2000; 2000US-0251030P.
PR   05-DEC-2000; 2000US-0251988P.
PR   05-DEC-2000; 2000US-0256719P.
PR   06-DEC-2000; 2000US-0251479P.
PR   08-DEC-2000; 2000US-0251856P.
PR   08-DEC-2000; 2000US-0251868P.
PR   08-DEC-2000; 2000US-0251869P.
PR   08-DEC-2000; 2000US-0251989P.
PR   08-DEC-2000; 2000US-0251990P.
PR   11-DEC-2000; 2000US-0254097P.
PR   05-JAN-2001; 2001US-0259678P.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Rosen CA,  Barash SC,  Ruben SM;
XX
DR   WPI; 2001-476161/51.
DR   N-PSDB; ABA06672.
XX
PT   Isolated nucleic acid molecule encoding an inflammation-associated
PT   polypeptide is used in preventing, treating or ameliorating a medical
PT   condition.
XX
PS   Claim 11; SEQ ID NO 758; 859pp + Sequence Listing; English.
XX
CC   The present invention provides human cDNAs, proteins and related genomic
CC   DNAs. These can be used in the treatment of neural, immune system,
CC   muscular, reproductive, gastrointestinal, pulmonary, cardiovascular,
CC   renal and proliferative disorders and inflammation. The present sequence
CC   is a protein of the invention
XX
SQ   Sequence 238 AA;

  Query Match             25.0%;  Score 278.5;  DB 4;  Length 238;
  Best Local Similarity   37.3%;  Pred. No. 1e-18;
  Matches   59;  Conservative   25;  Mismatches   65;  Indels    9;  Gaps    2;

Qy         51 LEVRTSTASGLLLWQGVEVGEAGQGKDFISLGLQDGHLVFRYQLGSGEARLVSEDPINDG 110
              :  :|:   |||||:|          |||||||:|| ||| | |||| | ::     |||
Db         87 MRFKTTAKDGLLLWRG--DSPMRPNSDFISLGLRDGALVFSYNLGSGVASIMVNGSFNDG 144

Qy        111 EWHRVTALREGRRGSIQVDGEELVSGRSPGPNVAVNAKGSVYIGGAPDVATLTGGRFSSG 170
               |||| |:|:|: | | ||     :|:|||    :|  |::|:||  ::|  |   :: |
Db        145 RWHRVKAVRDGQSGKITVDDYGARTGKSPGMMRQLNINGALYVGGMKEIALHTNRQYMRG 204

Qy        171 ITGCVKNLVLHSARPGAPPPQPLDLQHRAQAGANTRPC 208
              : ||: :  | :          : |   |  |  |   |
Db        205 LVGCISHFTLST-------DYHISLVEDAVDGKNINTC 235

RESULT 14
AAU19961
ID   AAU19961 standard; protein; 238 AA.
XX
AC   AAU19961;
XX
DT   06-DEC-2001  (first entry)
XX
DE   Novel human calcium-binding protein #70.
XX
KW   Human; calcium-binding protein; calcium flux; neurological disease;
KW   immune dysfunction; digestive disorder; neoplastic disease;
KW   blood disorder; infectious disease; gene therapy; immunosuppressive;
KW   antiarthritic; cytostatic; vasotropic; antibacterial; nootropic;
KW   virucide.
XX
```

```
OS   Homo sapiens.
XX
PN   WO200155304-A2.
XX
PD   02-AUG-2001.
XX
PF   17-JAN-2001; 2001WO-US001302.
XX
PR   31-JAN-2000; 2000US-0179065P.
PR   04-FEB-2000; 2000US-0180628P.
PR   24-FEB-2000; 2000US-0184664P.
PR   02-MAR-2000; 2000US-0186350P.
PR   16-MAR-2000; 2000US-0189874P.
PR   17-MAR-2000; 2000US-0190076P.
PR   18-APR-2000; 2000US-0198123P.
PR   19-MAY-2000; 2000US-0205515P.
PR   07-JUN-2000; 2000US-0209467P.
PR   28-JUN-2000; 2000US-0214886P.
PR   30-JUN-2000; 2000US-0215135P.
PR   07-JUL-2000; 2000US-0216647P.
PR   07-JUL-2000; 2000US-0216880P.
PR   11-JUL-2000; 2000US-0217487P.
PR   11-JUL-2000; 2000US-0217496P.
PR   14-JUL-2000; 2000US-0218290P.
PR   26-JUL-2000; 2000US-0220963P.
PR   26-JUL-2000; 2000US-0220964P.
PR   14-AUG-2000; 2000US-0224518P.
PR   14-AUG-2000; 2000US-0224519P.
PR   14-AUG-2000; 2000US-0225213P.
PR   14-AUG-2000; 2000US-0225214P.
PR   14-AUG-2000; 2000US-0225266P.
PR   14-AUG-2000; 2000US-0225267P.
PR   14-AUG-2000; 2000US-0225268P.
PR   14-AUG-2000; 2000US-0225270P.
PR   14-AUG-2000; 2000US-0225447P.
PR   14-AUG-2000; 2000US-0225757P.
PR   14-AUG-2000; 2000US-0225758P.
PR   14-AUG-2000; 2000US-0225759P.
PR   18-AUG-2000; 2000US-0226279P.
PR   22-AUG-2000; 2000US-0226681P.
PR   22-AUG-2000; 2000US-0226868P.
PR   22-AUG-2000; 2000US-0227182P.
PR   23-AUG-2000; 2000US-0227009P.
PR   30-AUG-2000; 2000US-0228924P.
PR   01-SEP-2000; 2000US-0229287P.
PR   01-SEP-2000; 2000US-0229343P.
PR   01-SEP-2000; 2000US-0229344P.
PR   01-SEP-2000; 2000US-0229345P.
PR   05-SEP-2000; 2000US-0229509P.
PR   05-SEP-2000; 2000US-0229513P.
PR   06-SEP-2000; 2000US-0230437P.
PR   06-SEP-2000; 2000US-0230438P.
PR   08-SEP-2000; 2000US-0231242P.
PR   08-SEP-2000; 2000US-0231243P.
PR   08-SEP-2000; 2000US-0231244P.
PR   08-SEP-2000; 2000US-0231413P.
PR   08-SEP-2000; 2000US-0231414P.
PR   08-SEP-2000; 2000US-0232080P.
PR   08-SEP-2000; 2000US-0232081P.
PR   12-SEP-2000; 2000US-0231968P.
PR   14-SEP-2000; 2000US-0232397P.
PR   14-SEP-2000; 2000US-0232398P.
PR   14-SEP-2000; 2000US-0232399P.
PR   14-SEP-2000; 2000US-0232400P.
PR   14-SEP-2000; 2000US-0232401P.
PR   14-SEP-2000; 2000US-0233063P.
PR   14-SEP-2000; 2000US-0233064P.
PR   14-SEP-2000; 2000US-0233065P.
PR   21-SEP-2000; 2000US-0234223P.
PR   21-SEP-2000; 2000US-0234274P.
PR   25-SEP-2000; 2000US-0234997P.
PR   25-SEP-2000; 2000US-0234998P.
PR   26-SEP-2000; 2000US-0235484P.
```

```
PR   27-SEP-2000; 2000US-0235834P.
PR   27-SEP-2000; 2000US-0235836P.
PR   29-SEP-2000; 2000US-0236327P.
PR   29-SEP-2000; 2000US-0236367P.
PR   29-SEP-2000; 2000US-0236368P.
PR   29-SEP-2000; 2000US-0236369P.
PR   29-SEP-2000; 2000US-0236370P.
PR   02-OCT-2000; 2000US-0236802P.
PR   02-OCT-2000; 2000US-0237037P.
PR   02-OCT-2000; 2000US-0237038P.
PR   02-OCT-2000; 2000US-0237039P.
PR   02-OCT-2000; 2000US-0237040P.
PR   13-OCT-2000; 2000US-0239935P.
PR   13-OCT-2000; 2000US-0239937P.
PR   20-OCT-2000; 2000US-0240960P.
PR   20-OCT-2000; 2000US-0241221P.
PR   20-OCT-2000; 2000US-0241785P.
PR   20-OCT-2000; 2000US-0241786P.
PR   20-OCT-2000; 2000US-0241787P.
PR   20-OCT-2000; 2000US-0241808P.
PR   20-OCT-2000; 2000US-0241809P.
PR   20-OCT-2000; 2000US-0241826P.
PR   01-NOV-2000; 2000US-0244617P.
PR   08-NOV-2000; 2000US-0246474P.
PR   08-NOV-2000; 2000US-0246475P.
PR   08-NOV-2000; 2000US-0246476P.
PR   08-NOV-2000; 2000US-0246477P.
PR   08-NOV-2000; 2000US-0246478P.
PR   08-NOV-2000; 2000US-0246523P.
PR   08-NOV-2000; 2000US-0246524P.
PR   08-NOV-2000; 2000US-0246525P.
PR   08-NOV-2000; 2000US-0246526P.
PR   08-NOV-2000; 2000US-0246527P.
PR   08-NOV-2000; 2000US-0246528P.
PR   08-NOV-2000; 2000US-0246532P.
PR   08-NOV-2000; 2000US-0246609P.
PR   08-NOV-2000; 2000US-0246610P.
PR   08-NOV-2000; 2000US-0246611P.
PR   08-NOV-2000; 2000US-0246613P.
PR   17-NOV-2000; 2000US-0249207P.
PR   17-NOV-2000; 2000US-0249208P.
PR   17-NOV-2000; 2000US-0249209P.
PR   17-NOV-2000; 2000US-0249210P.
PR   17-NOV-2000; 2000US-0249211P.
PR   17-NOV-2000; 2000US-0249212P.
PR   17-NOV-2000; 2000US-0249213P.
PR   17-NOV-2000; 2000US-0249214P.
PR   17-NOV-2000; 2000US-0249215P.
PR   17-NOV-2000; 2000US-0249216P.
PR   17-NOV-2000; 2000US-0249217P.
PR   17-NOV-2000; 2000US-0249218P.
PR   17-NOV-2000; 2000US-0249244P.
PR   17-NOV-2000; 2000US-0249245P.
PR   17-NOV-2000; 2000US-0249264P.
PR   17-NOV-2000; 2000US-0249265P.
PR   17-NOV-2000; 2000US-0249297P.
PR   17-NOV-2000; 2000US-0249299P.
PR   17-NOV-2000; 2000US-0249300P.
PR   01-DEC-2000; 2000US-0250160P.
PR   01-DEC-2000; 2000US-0250391P.
PR   05-DEC-2000; 2000US-0251030P.
PR   05-DEC-2000; 2000US-0251988P.
PR   05-DEC-2000; 2000US-0256719P.
PR   06-DEC-2000; 2000US-0251479P.
PR   08-DEC-2000; 2000US-0251856P.
PR   08-DEC-2000; 2000US-0251868P.
PR   08-DEC-2000; 2000US-0251869P.
PR   08-DEC-2000; 2000US-0251989P.
PR   08-DEC-2000; 2000US-0251990P.
PR   11-DEC-2000; 2000US-0254097P.
PR   05-JAN-2001; 2001US-0259678P.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
```

```
XX
PI   Rosen CA,  Barash SC,  Ruben SM;
XX
DR   WPI; 2001-465568/50.
DR   N-PSDB; AAS31646.
XX
PT   Isolated nucleic acid molecule encoding a calcium-binding protein is used
PT   in preventing, treating or ameliorating a medical condition.
XX
PS   Claim 11; SEQ ID NO 158; 542pp; English.
XX
CC   The present invention relates to the isolation of novel human calcium-
CC   binding proteins, and cDNA (AAS31577-AAS31654) and genomic sequences
CC   encoding for these proteins. The sequences of the invention are useful in
CC   the diagnosis, prevention and/or prognosis of diseases associated with
CC   aberrant calcium flux. Such disorders include neurological diseases (e.g.
CC   amylotrophic lateral sclerosis, ALS), immune dysfunction (e.g. severe
CC   combined immunodeficiency, SCID), digestive disorders (e.g. irritable
CC   bowel syndrome, IBS), neoplastic disease (e.g. cancer), blood disorders
CC   (e.g. haemophilia), and/or infectious disease (e.g. acquired
CC   immunodeficiency syndrome, AIDS). The novel calcium-binding proteins are
CC   also useful as screening tools to identify antagonists and/or agonists
CC   that may enhance or inhibit activities mediated by calcium-binding
CC   proteins. The polynucleotides of the invention are also useful in gene
CC   therapy. AAU19892-AAU19969 represent the novel human calcium-binding
CC   proteins. Note: The sequence data for this patent did not form part of
CC   the printed specification, but was obtained in electronic format directly
CC   from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 238 AA;

  Query Match             25.0%;  Score 278.5;  DB 4;  Length 238;
  Best Local Similarity   37.3%;  Pred. No. 1e-18;
  Matches   59;  Conservative   25;  Mismatches   65;  Indels    9;  Gaps    2;

Qy          51 LEVRTSTASGLLLWQGVEVGEAGQGKDFISLGLQDGHLVFRYQLGSGEARLVSEDPINDG 110
               :  :|:   |||||:|          |||||||:|| ||| | |||| | ::     |||
Db          87 MRFKTTAKDGLLLWRG--DSPMRPNSDFISLGLRDGALVFSYNLGSGVASIMVNGSFNDG 144

Qy         111 EWHRVTALREGRRGSIQVDGEELVSGRSPGPNVAVNAKGSVYIGGAPDVATLTGGRFSSG 170
                |||| |:|:|: | | ||     :|:|||    :|  |::|:||  ::|  |    ::|
Db         145 RWHRVKAVRDGQSGKITVDDYGARTGKSPGMMRQLNINGALYVGGMKEIALHTNRQYMRG 204

Qy         171 ITGCVKNLVLHSARPGAPPPQPLDLQHRAQAGANTRPC 208
               : ||: :  | :          : |   |  | |   |
Db         205 LVGCISHFTLST-------DYHISLVEDAVDGKNINTC 235


RESULT 15
ABJ05772
ID   ABJ05772 standard; protein; 238 AA.
XX
AC   ABJ05772;
XX
DT   14-NOV-2002  (first entry)
XX
DE   Novel human protein SEQ ID No 121.
XX
KW   Immunostimulant; antirheumatic; antiarthritic; neuroprotective;
KW   antiallergic; antidiabetic; antiasthmatic; antiinflammatory; nootropic;
KW   immunosuppressive; anticoagulant; thrombolytic; antiatherosclerotic;
KW   cytostatic; nephrotropic; antiparkinsonian; gynecological; virucide;
KW   antibacterial; antiarrhythmic; fungicide; HCFAT05; HWAAE95; HTNBM01;
KW   immunodeficiency; autoimmune disorder; allergic reaction; cardiovascular;
KW   inflammatory condition; graft-versus-host disease; reproductive system;
KW   blood-related disorder; hyperproliferative; endocrine; neurological;
KW   respiratory; renal; infectious disease; gastrointestinal; gene therapy;
KW   neuronal growth; neuronal disorder; neuro-degenerative condition;
KW   keratinocyte growth; human.
XX
OS   Homo sapiens.
XX
PN   US2002086330-A1.
```

```
XX
PD   04-JUL-2002.
XX
PF   17-JAN-2001; 2001US-00764893.
XX
PR   31-JAN-2000; 2000US-0179065P.
PR   04-FEB-2000; 2000US-0180628P.
PR   28-JUN-2000; 2000US-0214886P.
PR   07-JUL-2000; 2000US-0216647P.
PR   07-JUL-2000; 2000US-0216880P.
PR   11-JUL-2000; 2000US-0217487P.
PR   11-JUL-2000; 2000US-0217496P.
PR   14-JUL-2000; 2000US-0218290P.
PR   26-JUL-2000; 2000US-0220963P.
PR   26-JUL-2000; 2000US-0220964P.
PR   14-AUG-2000; 2000US-0224518P.
PR   14-AUG-2000; 2000US-0224519P.
PR   14-AUG-2000; 2000US-0225267P.
PR   14-AUG-2000; 2000US-0225268P.
PR   14-AUG-2000; 2000US-0225270P.
PR   14-AUG-2000; 2000US-0225447P.
PR   14-AUG-2000; 2000US-0225757P.
PR   14-AUG-2000; 2000US-0225758P.
PR   22-AUG-2000; 2000US-0226868P.
PR   30-AUG-2000; 2000US-0228924P.
PR   01-SEP-2000; 2000US-0229287P.
PR   01-SEP-2000; 2000US-0229343P.
PR   01-SEP-2000; 2000US-0229344P.
PR   01-SEP-2000; 2000US-0229345P.
PR   05-SEP-2000; 2000US-0229509P.
PR   05-SEP-2000; 2000US-0229513P.
PR   08-SEP-2000; 2000US-0231413P.
PR   21-SEP-2000; 2000US-0234223P.
PR   21-SEP-2000; 2000US-0234274P.
PR   25-SEP-2000; 2000US-0234997P.
PR   27-SEP-2000; 2000US-0235834P.
PR   29-SEP-2000; 2000US-0236327P.
PR   29-SEP-2000; 2000US-0236367P.
PR   29-SEP-2000; 2000US-0236368P.
PR   29-SEP-2000; 2000US-0236369P.
PR   29-SEP-2000; 2000US-0236370P.
PR   02-OCT-2000; 2000US-0236802P.
PR   02-OCT-2000; 2000US-0237037P.
PR   02-OCT-2000; 2000US-0237038P.
PR   02-OCT-2000; 2000US-0237039P.
PR   02-OCT-2000; 2000US-0237040P.
PR   13-OCT-2000; 2000US-0239935P.
PR   20-OCT-2000; 2000US-0240960P.
PR   20-OCT-2000; 2000US-0241785P.
PR   20-OCT-2000; 2000US-0241809P.
PR   01-NOV-2000; 2000US-0244617P.
PR   17-NOV-2000; 2000US-0249299P.
PR   08-DEC-2000; 2000US-0251856P.
PR   08-DEC-2000; 2000US-0251868P.
PR   08-DEC-2000; 2000US-0251869P.
XX
PA   (ROSE/) ROSEN C A.
PA   (RUBE/) RUBEN S M.
PA   (BARA/) BARASH S C.
XX
PI   Rosen CA,  Ruben SM,  Barash SC;
XX
DR   WPI; 2002-665432/71.
XX
PT   Novel polypeptide useful for diagnosis, prognosis, prevention, and
PT   treatment of immune, hyperproliferative, renal, respiratory,
PT   cardiovascular, reproductive, endocrine, gastrointestinal and
PT   neurological disorders.
XX
PS   Disclosure; Page 287-288; 335pp; English.
XX
CC   The invention relates to an isolated polypeptide comprising a sequence at
CC   least 90% identical to a full length protein sequence selected from 55
CC   sequences given in the specification such as a sequence of 163, 74 or 140
CC   amino acids fully defined in the specification, or the encoding sequence
CC   contained in 49 cDNA clones given in specification e.g. HCFAT05, HWAAE95
CC   or HTNBM01. The protein and its encoding nucleic acid are useful for
CC   diagnosing a pathological condition or susceptibility to a pathological
CC   condition in a subject and for preventing, treating or ameliorating a
CC   medical condition. The protein, its encoding nucleic acid and an isolated
CC   antibody that can bind to the protein are useful in treating, preventing,
CC   diagnosing and/or prognosing immunodeficiencies, autoimmune disorders,
CC   allergic reactions and conditions, inflammatory conditions, graft-versus-
CC   host disease, blood-related disorders, hyperproliferative disorders,
CC   renal disorders, cardiovascular disorders, respiratory disorders,
CC   neurological disorders, endocrine disorders, reproductive system
CC   disorders, infectious diseases, and gastrointestinal disorders. The
CC   protein of the invention is useful to stimulate neuronal growth and to
CC   treat, prevent, and/or diagnose neuronal damage which occurs in certain
CC   neuronal disorders or neuro-degenerative conditions, for stimulating
CC   keratinocyte growth, to prevent hair loss, to modulate mammalian
CC   characteristics such as body height, weight, hair color, and to increase
CC   or decrease storage capabilities, fat content, lipid, protein,
CC   carbohydrate, vitamins, minerals, cofactors or other nutritional
CC   components. The nucleic acid of the invention can be used in gene
CC   therapy. This sequence represents a novel human protein of the invention
XX
SQ   Sequence 238 AA;

  Query Match             25.0%;  Score 278.5;  DB 5;  Length 238;
  Best Local Similarity   37.3%;  Pred. No. 1e-18;
  Matches   59;  Conservative   25;  Mismatches   65;  Indels    9;  Gaps    2;

Qy         51 LEVRTSTASGLLLWQGVEVGEAGQGKDFISLGLQDGHLVFRYQLGSGEARLVSEDPINDG 110
              :  :|:   |||||:|          |||||||:|| ||| | |||| | ::     |||
Db         87 MRFKTTAKDGLLLWRG--DSPMRPNSDFISLGLRDGALVFSYNLGSGVASIMVNGSFNDG 144

Qy        111 EWHRVTALREGRRGSIQVDGEELVSGRSPGPNVAVNAKGSVYIGGAPDVATLTGGRFSSG 170
               |||| |:|:|: | | ||     :|:|||    :|  |::|:||  ::|  |  ::  |
Db        145 RWHRVKAVRDGQSGKITVDDYGARTGKSPGMMRQLNINGALYVGGMKEIALHTNRQYMRG 204

Qy        171 ITGCVKNLVLHSARPGAPPPQPLDLQHRAQAGANTRPC 208
              : ||: :  | :          : |   |  | |   |
Db        205 LVGCISHFTLST-------DYHISLVEDAVDGKNINTC 235

Search completed: March  9, 2004, 17:19:04
Job time : 26.5713 secs
```

```
                        GenCore version 5.1.6
              Copyright (c) 1993 - 2004  Compugen Ltd.

OM protein - protein search, using sw model

Run on:          March  9, 2004, 17:16:23 ; Search time 8.23186 Seconds
                                            (without alignments)
                                            1317.011 Million cell updates/sec

Title:           US-10-006-011A-10
Perfect score:   1115
Sequence:        1 GIAESDWHLEGSGGNDAPGQ..........QPLDLQHRAQAGANTRPCPS 210

Scoring table:   BLOSUM62
                 Gapop 10.0 , Gapext 0.5

Searched:        389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters:      389414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :        Issued_Patents_AA:*
                 1:  /cgn2_6/ptodata/2/iaa/5A_COMB.pep:*
                 2:  /cgn2_6/ptodata/2/iaa/5B_COMB.pep:*
                 3:  /cgn2_6/ptodata/2/iaa/6A_COMB.pep:*
                 4:  /cgn2_6/ptodata/2/iaa/6B_COMB.pep:*
                 5:  /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep:*
                 6:  /cgn2_6/ptodata/2/iaa/backfiles1.pep:*

      Pred. No. is the number of results predicted by chance to have a
      score greater than or equal to the score of the result being printed,
      and is derived by analysis of the total score distribution.

                              SUMMARIES
                %
Result          Query
  No.    Score  Match Length DB  ID                         Description
-------------------------------------------------------------------------------
    1      261   23.4    216  4  US-09-077-955-31             Sequence 31, Appl
    2      261   23.4    256  4  US-09-077-955-30             Sequence 30, Appl
    3      261   23.4    294  4  US-09-077-955-29             Sequence 29, Appl
    4      261   23.4    338  4  US-09-077-955-28             Sequence 28, Appl
    5      261   23.4    390  4  US-09-077-955-27             Sequence 27, Appl
    6      261   23.4    440  4  US-09-077-955-26             Sequence 26, Appl
    7      261   23.4    456  4  US-09-077-955-25             Sequence 25, Appl
    8      261   23.4    492  2  US-08-644-271-32             Sequence 32, Appl
    9      261   23.4    492  4  US-09-077-955-36             Sequence 36, Appl
   10    242.5   21.7   1940  2  US-08-644-271-30             Sequence 30, Appl
   11    242.5   21.7   1940  4  US-09-077-955-34             Sequence 34, Appl
   12    223.5   20.0   1130  2  US-08-460-309-2              Sequence 2, Appli
   13    223.5   20.0   1130  2  US-08-125-077-2              Sequence 2, Appli
   14    223.5   20.0   1130  6  5444158-2                   Patent No. 5444158
   15    223.5   20.0   3088  4  US-09-562-702A-8             Sequence 8, Appli
   16    223.5   20.0   3089  4  US-09-562-702A-4             Sequence 4, Appli
   17    223.5   20.0   3110  4  US-09-562-702A-2             Sequence 2, Appli
   18    223.5   20.0   3110  4  US-09-562-702A-6             Sequence 6, Appli
   19    223.5   20.0   3110  4  US-09-561-709B-7             Sequence 7, Appli
   20    223.5   20.0   3111  2  US-08-460-309-4              Sequence 4, Appli
   21    223.5   20.0   3111  2  US-08-125-077-4              Sequence 4, Appli
   22    207.5   18.6   3084  4  US-09-562-702A-12            Sequence 12, Appl
   23    207.5   18.6   3106  4  US-09-562-702A-10            Sequence 10, Appl
   24    188.5   16.9   3075  2  US-08-460-309-5              Sequence 5, Appli
   25    188.5   16.9   3075  2  US-08-125-077-5              Sequence 5, Appli
   26    173.5   15.6   1792  4  US-09-561-818A-12            Sequence 12, Appl
   27    173.5   15.6   1816  4  US-09-561-818A-10            Sequence 10, Appl
   28    165.5   14.8   1792  4  US-09-561-818A-4             Sequence 4, Appli
   29    165.5   14.8   1800  4  US-09-561-818A-8             Sequence 8, Appli
   30    165.5   14.8   1816  4  US-09-561-818A-2             Sequence 2, Appli
   31    165.5   14.8   1824  4  US-09-561-818A-6             Sequence 6, Appli
   32    162.5   14.6   1713  3  US-08-600-982-24             Sequence 24, Appl
   33    162.5   14.6   1713  5  PCT-US94-10261A-24           Sequence 24, Appl
   34      151   13.5   1525  3  US-09-191-647-2              Sequence 2, Appli
   35      151   13.5   1525  3  US-09-540-245A-2             Sequence 2, Appli
   36      151   13.5   1525  3  US-09-540-153-2              Sequence 2, Appli
   37      148   13.3   1529  4  US-09-312-283C-396           Sequence 396, App
   38    139.5   12.5    735  3  US-09-191-647-9              Sequence 9, Appli
   39    139.5   12.5    735  3  US-09-540-245A-9             Sequence 9, Appli
   40    139.5   12.5    735  3  US-09-540-153-9              Sequence 9, Appli
   41    138.5   12.4    953  4  US-09-845-583A-4             Sequence 4, Appli
   42      132   11.8    716  4  US-09-312-283C-183           Sequence 183, App
   43      132   11.8    771  3  US-09-188-930-183            Sequence 183, App
   44    130.5   11.7   1381  4  US-08-826-134-4              Sequence 4, Appli
   45    129.5   11.6   1384  4  US-08-826-134-2              Sequence 2, Appli


                              ALIGNMENTS


RESULT 1
US-09-077-955-31
; Sequence 31, Application US/09077955A
; Patent No. 6413740
; GENERAL INFORMATION:
;  APPLICANT: Valenzuela et al., David M.
;  TITLE OF INVENTION: NOVEL TYROSINE KINASE RECEPTORS AND LIGANDS
;  FILE REFERENCE: REG195-B-PCT-US
;  CURRENT APPLICATION NUMBER: US/09/077,955A
;  CURRENT FILING DATE: 1998-09-10
;  EARLIER APPLICATION NUMBER: PCT/US96/20696
;  EARLIER FILING DATE: 1996-12-13
;  EARLIER APPLICATION NUMBER: 08/644,271
;  EARLIER FILING DATE: 1996-05-10
;  EARLIER APPLICATION NUMBER: 60/008,657
;  EARLIER FILING DATE: 1995-12-15
;  NUMBER OF SEQ ID NOS: 36
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 31
;   LENGTH: 216
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-077-955-31

  Query Match             23.4%;  Score 261;  DB 4;  Length 216;
  Best Local Similarity   32.8%;  Pred. No. 1.7e-18;
  Matches   65;  Conservative   27;  Mismatches   76;  Indels   30;  Gaps    4;

Qy          31 LAFPGHVFSRSLPEVPET----------------IELEVRTSTASGLLLWQGVEVGEAGQ 74
               ||| |  |   |  | |:                 || :||    ||:||     |:| :
Db          30 LAFDGRTFVEYLNAVTESELANEIPVEKALQSNHFELSLRTEATQGLVLWS----GKATE 85

Qy          75 GKDFISLGLQDGHLVFRYQLGSGEARLVSEDPINDGEWHRVTALREGRRGSIQVDGEELV 134
                 |:::| : ||||   | |||    | |  |:|   | || | || | ||:||  |  |
Db          86 RADYVALAIVDGHLQLSYNLGSQPVVLRSTVPVNTNRWLRVVAHREQREGSLQVGNEAPV 145

Qy         135 SGRSPGPNVAVNAKGSVYIGGAPD--VATLTGGRFSSGITGCVKNLVLHSARPGAPPPQP 192
               :| ||     ::  |::::|| |:  |       : :|  ||::::|:           |
Db         146 TGSSPLGATQLDTDGALWLGGLPELPVGPALPKAYGTGFVGCLRDVVVGR--------HP 197

Qy         193 LDLQHRAQAGANTRPCPS 210
               | |   |      ||||:
Db         198 LHLLEDAVTKPELRPCPT 215


RESULT 2
US-09-077-955-30
; Sequence 30, Application US/09077955A
; Patent No. 6413740
; GENERAL INFORMATION:
```

```
;  APPLICANT: Valenzuela et al., David M.
;  TITLE OF INVENTION: NOVEL TYROSINE KINASE RECEPTORS AND LIGANDS
;  FILE REFERENCE: REG195-B-PCT-US
;  CURRENT APPLICATION NUMBER: US/09/077,955A
;  CURRENT FILING DATE: 1998-09-10
;  EARLIER APPLICATION NUMBER: PCT/US96/20696
;  EARLIER FILING DATE: 1996-12-13
;  EARLIER APPLICATION NUMBER: 08/644,271
;  EARLIER FILING DATE: 1996-05-10
;  EARLIER APPLICATION NUMBER: 60/008,657
;  EARLIER FILING DATE: 1995-12-15
;  NUMBER OF SEQ ID NOS: 36
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 30
;   LENGTH: 256
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-077-955-30

  Query Match             23.4%;  Score 261;  DB 4;  Length 256;
  Best Local Similarity   32.8%;  Pred. No. 2.2e-18;
  Matches   65;  Conservative   27;  Mismatches   76;  Indels   30;  Gaps    4;

Qy          31 LAFPGHVFSRSLPEVPET----------------IELEVRTSTASGLLLWQGVEVGEAGQ 74
               ||| |  |   |  | |:                 || :||    ||:||     |:| :
Db          70 LAFDGRTFVEYLNAVTESELANEIPVEKALQSNHFELSLRTEATQGLVLWS----GKATE 125

Qy          75 GKDFISLGLQDGHLVFRYQLGSGEARLVSEDPINDGEWHRVTALREGRRGSIQVDGEELV 134
                 |:::| : ||||   | |||    | |  |:|   | || | || | ||:||  |  |
Db         126 RADYVALAIVDGHLQLSYNLGSQPVVLRSTVPVNTNRWLRVVAHREQREGSLQVGNEAPV 185

Qy         135 SGRSPGPNVAVNAKGSVYIGGAPD--VATLTGGRFSSGITGCVKNLVLHSARPGAPPPQP 192
               :| ||     ::  |::::|| |:  |       : :|  ||::::|:           |
Db         186 TGSSPLGATQLDTDGALWLGGLPELPVGPALPKAYGTGFVGCLRDVVVGR--------HP 237

Qy         193 LDLQHRAQAGANTRPCPS 210
               | |   |      ||||:
Db         238 LHLLEDAVTKPELRPCPT 255

RESULT 3
US-09-077-955-29
; Sequence 29, Application US/09077955A
; Patent No. 6413740
; GENERAL INFORMATION:
;  APPLICANT: Valenzuela et al., David M.
;  TITLE OF INVENTION: NOVEL TYROSINE KINASE RECEPTORS AND LIGANDS
;  FILE REFERENCE: REG195-B-PCT-US
;  CURRENT APPLICATION NUMBER: US/09/077,955A
;  CURRENT FILING DATE: 1998-09-10
;  EARLIER APPLICATION NUMBER: PCT/US96/20696
;  EARLIER FILING DATE: 1996-12-13
;  EARLIER APPLICATION NUMBER: 08/644,271
;  EARLIER FILING DATE: 1996-05-10
;  EARLIER APPLICATION NUMBER: 60/008,657
;  EARLIER FILING DATE: 1995-12-15
;  NUMBER OF SEQ ID NOS: 36
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 29
;   LENGTH: 294
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-077-955-29

  Query Match             23.4%;  Score 261;  DB 4;  Length 294;
  Best Local Similarity   32.8%;  Pred. No. 2.6e-18;
  Matches   65;  Conservative   27;  Mismatches   76;  Indels   30;  Gaps    4;

Qy          31 LAFPGHVFSRSLPEVPET----------------IELEVRTSTASGLLLWQGVEVGEAGQ 74
               ||| |  |   |  | |:                 || :||    ||:||     |:| :
Db         108 LAFDGRTFVEYLNAVTESELANEIPVEKALQSNHFELSLRTEATQGLVLWS----GKATE 163

Qy          75 GKDFISLGLQDGHLVFRYQLGSGEARLVSEDPINDGEWHRVTALREGRRGSIQVDGEELV 134
```

```
                 |:::| : ||||   | |||    | |  |:|   | || | || | ||:||  |  |
Db         164 RADYVALAIVDGHLQLSYNLGSQPVVLRSTVPVNTNRWLRVVAHREQREGSLQVGNEAPV 223

Qy         135 SGRSPGPNVAVNAKGSVYIGGAPD--VATLTGGRFSSGITGCVKNLVLHSARPGAPPPQP 192
               :| ||     ::  |::::|| |:  |       : :|  ||::::|:           |
Db         224 TGSSPLGATQLDTDGALWLGGLPELPVGPALPKAYGTGFVGCLRDVVVGR--------HP 275

Qy         193 LDLQHRAQAGANTRPCPS 210
               | |   |      ||||:
Db         276 LHLLEDAVTKPELRPCPT 293

RESULT 4
US-09-077-955-28
; Sequence 28, Application US/09077955A
; Patent No. 6413740
; GENERAL INFORMATION:
;  APPLICANT: Valenzuela et al., David M.
;  TITLE OF INVENTION: NOVEL TYROSINE KINASE RECEPTORS AND LIGANDS
;  FILE REFERENCE: REG195-B-PCT-US
;  CURRENT APPLICATION NUMBER: US/09/077,955A
;  CURRENT FILING DATE: 1998-09-10
;  EARLIER APPLICATION NUMBER: PCT/US96/20696
;  EARLIER FILING DATE: 1996-12-13
;  EARLIER APPLICATION NUMBER: 08/644,271
;  EARLIER FILING DATE: 1996-05-10
;  EARLIER APPLICATION NUMBER: 60/008,657
;  EARLIER FILING DATE: 1995-12-15
;  NUMBER OF SEQ ID NOS: 36
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 28
;   LENGTH: 338
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-077-955-28

  Query Match             23.4%;  Score 261;  DB 4;  Length 338;
  Best Local Similarity   32.8%;  Pred. No. 3.2e-18;
  Matches   65;  Conservative   27;  Mismatches   76;  Indels   30;  Gaps    4;

Qy          31 LAFPGHVFSRSLPEVPET----------------IELEVRTSTASGLLLWQGVEVGEAGQ 74
               ||| |  |   |  | |:                 || :||    ||:||     |:| :
Db         152 LAFDGRTFVEYLNAVTESELANEIPVEKALQSNHFELSLRTEATQGLVLWS----GKATE 207

Qy          75 GKDFISLGLQDGHLVFRYQLGSGEARLVSEDPINDGEWHRVTALREGRRGSIQVDGEELV 134
                 |:::| : ||||   | |||    | |  |:|   | || | || | ||:||  |  |
Db         208 RADYVALAIVDGHLQLSYNLGSQPVVLRSTVPVNTNRWLRVVAHREQREGSLQVGNEAPV 267

Qy         135 SGRSPGPNVAVNAKGSVYIGGAPD--VATLTGGRFSSGITGCVKNLVLHSARPGAPPPQP 192
               :| ||     ::  |::::|| |:  |       : :|  ||::::|:           |
Db         268 TGSSPLGATQLDTDGALWLGGLPELPVGPALPKAYGTGFVGCLRDVVVGR--------HP 319

Qy         193 LDLQHRAQAGANTRPCPS 210
               | |   |      ||||:
Db         320 LHLLEDAVTKPELRPCPT 337

RESULT 5
US-09-077-955-27
; Sequence 27, Application US/09077955A
; Patent No. 6413740
; GENERAL INFORMATION:
;  APPLICANT: Valenzuela et al., David M.
;  TITLE OF INVENTION: NOVEL TYROSINE KINASE RECEPTORS AND LIGANDS
;  FILE REFERENCE: REG195-B-PCT-US
;  CURRENT APPLICATION NUMBER: US/09/077,955A
;  CURRENT FILING DATE: 1998-09-10
;  EARLIER APPLICATION NUMBER: PCT/US96/20696
;  EARLIER FILING DATE: 1996-12-13
;  EARLIER APPLICATION NUMBER: 08/644,271
;  EARLIER FILING DATE: 1996-05-10
;  EARLIER APPLICATION NUMBER: 60/008,657
;  EARLIER FILING DATE: 1995-12-15
```

```
;  NUMBER OF SEQ ID NOS: 36
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 27
;   LENGTH: 390
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-077-955-27

  Query Match             23.4%;  Score 261;  DB 4;  Length 390;
  Best Local Similarity   32.8%;  Pred. No. 3.9e-18;
  Matches   65;  Conservative   27;  Mismatches   76;  Indels   30;  Gaps    4;

Qy          31 LAFPGHVFSRSLPEVPET----------------IELEVRTSTASGLLLWQGVEVGEAGQ 74
               ||| |  |   |  | |:                 || :||    ||:||     |:| :
Db         204 LAFDGRTFVEYLNAVTESELANEIPVEKALQSNHFELSLRTEATQGLVLWS----GKATE 259

Qy          75 GKDFISLGLQDGHLVFRYQLGSGEARLVSEDPINDGEWHRVTALREGRRGSIQVDGEELV 134
                 |:::| : ||||   | |||    | |  |:|   | || | || | ||:||  |  |
Db         260 RADYVALAIVDGHLQLSYNLGSQPVVLRSTVPVNTNRWLRVVAHREQREGSLQVGNEAPV 319

Qy         135 SGRSPGPNVAVNAKGSVYIGGAPD--VATLTGGRFSSGITGCVKNLVLHSARPGAPPPQP 192
               :| ||     ::  |::::|| |:  |       : :|  ||::::|:           |
Db         320 TGSSPLGATQLDTDGALWLGGLPELPVGPALPKAYGTGFVGCLRDVVVGR--------HP 371

Qy         193 LDLQHRAQAGANTRPCPS 210
               | |   |      ||||:
Db         372 LHLLEDAVTKPELRPCPT 389

RESULT 6
US-09-077-955-26
; Sequence 26, Application US/09077955A
; Patent No. 6413740
; GENERAL INFORMATION:
;  APPLICANT: Valenzuela et al., David M.
;  TITLE OF INVENTION: NOVEL TYROSINE KINASE RECEPTORS AND LIGANDS
;  FILE REFERENCE: REG195-B-PCT-US
;  CURRENT APPLICATION NUMBER: US/09/077,955A
;  CURRENT FILING DATE: 1998-09-10
;  EARLIER APPLICATION NUMBER: PCT/US96/20696
;  EARLIER FILING DATE: 1996-12-13
;  EARLIER APPLICATION NUMBER: 08/644,271
;  EARLIER FILING DATE: 1996-05-10
;  EARLIER APPLICATION NUMBER: 60/008,657
;  EARLIER FILING DATE: 1995-12-15
;  NUMBER OF SEQ ID NOS: 36
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 26
;   LENGTH: 440
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-077-955-26

  Query Match             23.4%;  Score 261;  DB 4;  Length 440;
  Best Local Similarity   32.8%;  Pred. No. 4.6e-18;
  Matches   65;  Conservative   27;  Mismatches   76;  Indels   30;  Gaps    4;

Qy          31 LAFPGHVFSRSLPEVPET----------------IELEVRTSTASGLLLWQGVEVGEAGQ 74
               ||| |  |   |  | |:                 || :||    ||:||     |:| :
Db         254 LAFDGRTFVEYLNAVTESELANEIPVEKALQSNHFELSLRTEATQGLVLWS----GKATE 309

Qy          75 GKDFISLGLQDGHLVFRYQLGSGEARLVSEDPINDGEWHRVTALREGRRGSIQVDGEELV 134
                 |:::| : ||||   | |||    | |  |:|   | || | || | ||:||  |  |
Db         310 RADYVALAIVDGHLQLSYNLGSQPVVLRSTVPVNTNRWLRVVAHREQREGSLQVGNEAPV 369

Qy         135 SGRSPGPNVAVNAKGSVYIGGAPD--VATLTGGRFSSGITGCVKNLVLHSARPGAPPPQP 192
               :| ||     ::  |::::|| |:  |       : :|  ||::::|:           |
Db         370 TGSSPLGATQLDTDGALWLGGLPELPVGPALPKAYGTGFVGCLRDVVVGR--------HP 421

Qy         193 LDLQHRAQAGANTRPCPS 210
               | |   |      ||||:
Db         422 LHLLEDAVTKPELRPCPT 439
```

```
RESULT 7
US-09-077-955-25
; Sequence 25, Application US/09077955A
; Patent No. 6413740
; GENERAL INFORMATION:
;  APPLICANT: Valenzuela et al., David M.
;  TITLE OF INVENTION: NOVEL TYROSINE KINASE RECEPTORS AND LIGANDS
;  FILE REFERENCE: REG195-B-PCT-US
;  CURRENT APPLICATION NUMBER: US/09/077,955A
;  CURRENT FILING DATE: 1998-09-10
;  EARLIER APPLICATION NUMBER: PCT/US96/20696
;  EARLIER FILING DATE: 1996-12-13
;  EARLIER APPLICATION NUMBER: 08/644,271
;  EARLIER FILING DATE: 1996-05-10
;  EARLIER APPLICATION NUMBER: 60/008,657
;  EARLIER FILING DATE: 1995-12-15
;  NUMBER OF SEQ ID NOS: 36
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 25
;   LENGTH: 456
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-077-955-25

  Query Match             23.4%;  Score 261;  DB 4;  Length 456;
  Best Local Similarity   32.8%;  Pred. No. 4.9e-18;
  Matches   65;  Conservative   27;  Mismatches   76;  Indels   30;  Gaps    4;

Qy          31 LAFPGHVFSRSLPEVPET----------------IELEVRTSTASGLLLWQGVEVGEAGQ 74
               ||| |  |   |  | |:                 || :||    ||:||     |:| :
Db         270 LAFDGRTFVEYLNAVTESELANEIPVEKALQSNHFELSLRTEATQGLVLWS----GKATE 325

Qy          75 GKDFISLGLQDGHLVFRYQLGSGEARLVSEDPINDGEWHRVTALREGRRGSIQVDGEELV 134
                 |:::| : ||||   | |||    | |  |:|   | || | || | ||:||  |  |
Db         326 RADYVALAIVDGHLQLSYNLGSQPVVLRSTVPVNTNRWLRVVAHREQREGSLQVGNEAPV 385

Qy         135 SGRSPGPNVAVNAKGSVYIGGAPD--VATLTGGRFSSGITGCVKNLVLHSARPGAPPPQP 192
               :| ||     ::  |::::|| |:  |       : :|  ||::::|:           |
Db         386 TGSSPLGATQLDTDGALWLGGLPELPVGPALPKAYGTGFVGCLRDVVVGR--------HP 437

Qy         193 LDLQHRAQAGANTRPCPS 210
               | |   |      ||||:
Db         438 LHLLEDAVTKPELRPCPT 455

RESULT 8
US-08-644-271-32
; Sequence 32, Application US/08644271
; Patent No. 5814478
;  GENERAL INFORMATION:
;    APPLICANT:  Valenzuela, et al.
;    TITLE OF INVENTION:  NOVEL TYROSINE KINASE RECEPTORS
;    TITLE OF INVENTION:  AND LIGANDS
;    NUMBER OF SEQUENCES:  32
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Regeneron Pharmaceuticals, Inc.
;      STREET:  777 Old Saw Mill Road
;      CITY:  Tarrytown
;      STATE:  NY
;      COUNTRY:  USA
;      ZIP:  10591
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/644,271
;      FILING DATE:  10-MAY-1996
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  USSN 60/008,657
```

```
;      FILING DATE:  15-DEC-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Cobert, Robert J
;      REGISTRATION NUMBER:  36,108
;      REFERENCE/DOCKET NUMBER:  REG 195A
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  914-345-7400
;      TELEFAX:  914-345-7721
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  32:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  492 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
;    FRAGMENT TYPE:  internal
;    FEATURE:
;      NAME/KEY:  Human Agrin
;      LOCATION:  1...492
;      OTHER INFORMATION:
US-08-644-271-32

  Query Match             23.4%;  Score 261;  DB 2;  Length 492;
  Best Local Similarity   32.8%;  Pred. No. 5.4e-18;
  Matches   65;  Conservative   27;  Mismatches   76;  Indels   30;  Gaps    4;

Qy           31 LAFPGHVFSRSLPEVPET----------------IELEVRTSTASGLLLWQGVEVGEAGQ 74
                ||| |  |   |  | |:                 || :||    ||:||     |:| :
Db          306 LAFDGRTFVEYLNAVTESELANEIPVEKALQSNHFELSLRTEATQGLVLWS----GKATE 361

Qy           75 GKDFISLGLQDGHLVFRYQLGSGEARLVSEDPINDGEWHRVTALREGRRGSIQVDGEELV 134
                  |:::| : ||||   | |||    | |  |:|   | || | || | ||:||  |  |
Db          362 RADYVALAIVDGHLQLSYNLGSQPVVLRSTVPVNTNRWLRVVAHREQREGSLQVGNEAPV 421

Qy          135 SGRSPGPNVAVNAKGSVYIGGAPD--VATLTGGRFSSGITGCVKNLVLHSARPGAPPPQP 192
                :| ||     ::  |::::|| |:  |       : :|  ||::::|:           |
Db          422 TGSSPLGATQLDTDGALWLGGLPELPVGPALPKAYGTGFVGCLRDVVVGR--------HP 473

Qy          193 LDLQHRAQAGANTRPCPS 210
                | |   |      ||||:
Db          474 LHLLEDAVTKPELRPCPT 491

RESULT 9
US-09-077-955-36
; Sequence 36, Application US/09077955A
; Patent No. 6413740
; GENERAL INFORMATION:
;  APPLICANT: Valenzuela et al., David M.
;  TITLE OF INVENTION: NOVEL TYROSINE KINASE RECEPTORS AND LIGANDS
;  FILE REFERENCE: REG195-B-PCT-US
;  CURRENT APPLICATION NUMBER: US/09/077,955A
;  CURRENT FILING DATE: 1998-09-10
;  EARLIER APPLICATION NUMBER: PCT/US96/20696
;  EARLIER FILING DATE: 1996-12-13
;  EARLIER APPLICATION NUMBER: 08/644,271
;  EARLIER FILING DATE: 1996-05-10
;  EARLIER APPLICATION NUMBER: 60/008,657
;  EARLIER FILING DATE: 1995-12-15
;  NUMBER OF SEQ ID NOS: 36
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 36
;   LENGTH: 492
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-077-955-36

  Query Match             23.4%;  Score 261;  DB 4;  Length 492;
  Best Local Similarity   32.8%;  Pred. No. 5.4e-18;
  Matches   65;  Conservative   27;  Mismatches   76;  Indels   30;  Gaps    4;

Qy           31 LAFPGHVFSRSLPEVPET----------------IELEVRTSTASGLLLWQGVEVGEAGQ 74
                ||| |  |   |  | |:                 || :||    ||:||     |:| :
Db          306 LAFDGRTFVEYLNAVTESELANEIPVEKALQSNHFELSLRTEATQGLVLWS----GKATE 361

Qy           75 GKDFISLGLQDGHLVFRYQLGSGEARLVSEDPINDGEWHRVTALREGRRGSIQVDGEELV 134
                  |:::| : ||||   | |||    | |  |:|   | || | || | ||:||  |  |
Db          362 RADYVALAIVDGHLQLSYNLGSQPVVLRSTVPVNTNRWLRVVAHREQREGSLQVGNEAPV 421

Qy          135 SGRSPGPNVAVNAKGSVYIGGAPD--VATLTGGRFSSGITGCVKNLVLHSARPGAPPPQP 192
                :| ||     ::  |::::|| |:  |       : :|  ||::::|:           |
Db          422 TGSSPLGATQLDTDGALWLGGLPELPVGPALPKAYGTGFVGCLRDVVVGR--------HP 473

Qy          193 LDLQHRAQAGANTRPCPS 210
                | |   |      ||||:
Db          474 LHLLEDAVTKPELRPCPT 491

RESULT 10
US-08-644-271-30
; Sequence 30, Application US/08644271
; Patent No. 5814478
;  GENERAL INFORMATION:
;    APPLICANT:  Valenzuela, et al.
;    TITLE OF INVENTION:  NOVEL TYROSINE KINASE RECEPTORS
;    TITLE OF INVENTION:  AND LIGANDS
;    NUMBER OF SEQUENCES:  32
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Regeneron Pharmaceuticals, Inc.
;      STREET:  777 Old Saw Mill Road
;      CITY:  Tarrytown
;      STATE:  NY
;      COUNTRY:  USA
;      ZIP:  10591
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/644,271
;      FILING DATE:  10-MAY-1996
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  USSN 60/008,657
;      FILING DATE:  15-DEC-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Cobert, Robert J
;      REGISTRATION NUMBER:  36,108
;      REFERENCE/DOCKET NUMBER:  REG 195A
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  914-345-7400
;      TELEFAX:  914-345-7721
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  30:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1940 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
;    FEATURE:
;      NAME/KEY:  Rat Agrin
;      LOCATION:  1...1940
;      OTHER INFORMATION:
US-08-644-271-30

  Query Match             21.7%;  Score 242.5;  DB 2;  Length 1940;
  Best Local Similarity   32.6%;  Pred. No. 2.7e-15;
  Matches   61;  Conservative   26;  Mismatches   87;  Indels   13;  Gaps    3;

Qy           25 FHDDGFLAFPGHVFSRSLPEVPETIELEVRTSTASGLLLWQGVEVGEAGQGKDFISLGLQ 84
                |    |||||      :|      : || |     ||||: |       :||||::| |
Db         1268 FKGHSFLAFPTLRAYHTL-----RLALEFRALETEGLLLYNG-----NARGKDFLALALL 1317
```

```
Qy          85 DGHLVFRYQLGSGEARLVSEDPINDGEWHRVTALREGRRGSIQVDGEELVSGRSPGPNVA 144
               || : ||:  ||| | | |  |:  | |||:   |  |:|:: ||||  | | ||
Db        1318 DGRVQFRFDTGSGPAVLTSLVPVEPGRWHRLELSRHWRQGTLSVDGETPVVGESPSGTDG 1377

Qy         145 VNAKGSVYIGGAPD---VATLTGGRFSSGITGCVKNLVLHSARPGAPPPQPLDLQHRAQA 201
               :|   ::|:|| |:      |       |: ||:: | :::  :     |   :|
Db        1378 LNLDTNLYVGGIPEEQVAMVLDRTSVGVGLKGCIRMLDINNQQLELSDWQRAAVQSSGVG 1437

Qy         202 GANTRPC 208
                    ||
Db        1438 ECGDHPC 1444

RESULT 11
US-09-077-955-34
; Sequence 34, Application US/09077955A
; Patent No. 6413740
; GENERAL INFORMATION:
;  APPLICANT: Valenzuela et al., David M.
;  TITLE OF INVENTION: NOVEL TYROSINE KINASE RECEPTORS AND LIGANDS
;  FILE REFERENCE: REG195-B-PCT-US
;  CURRENT APPLICATION NUMBER: US/09/077,955A
;  CURRENT FILING DATE: 1998-09-10
;  EARLIER APPLICATION NUMBER: PCT/US96/20696
;  EARLIER FILING DATE: 1996-12-13
;  EARLIER APPLICATION NUMBER: 08/644,271
;  EARLIER FILING DATE: 1996-05-10
;  EARLIER APPLICATION NUMBER: 60/008,657
;  EARLIER FILING DATE: 1995-12-15
;  NUMBER OF SEQ ID NOS: 36
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 34
;   LENGTH: 1940
;   TYPE: PRT
;   ORGANISM: Rattus sp.
US-09-077-955-34

  Query Match             21.7%;  Score 242.5;  DB 4;  Length 1940;
  Best Local Similarity   32.6%;  Pred. No. 2.7e-15;
  Matches   61;  Conservative   26;  Mismatches   87;  Indels   13;  Gaps    3;

Qy          25 FHDDGFLAFPGHVFSRSLPEVPETIELEVRTSTASGLLLWQGVEVGEAGQGKDFISLGLQ 84
               |    |||||      :|      : || |     ||||: |       :||||::| |
Db        1268 FKGHSFLAFPTLRAYHTL-----RLALEFRALETEGLLLYNG-----NARGKDFLALALL 1317

Qy          85 DGHLVFRYQLGSGEARLVSEDPINDGEWHRVTALREGRRGSIQVDGEELVSGRSPGPNVA 144
               || : ||:  ||| | | |  |:  | |||:   |  |:|:: ||||  | | ||
Db        1318 DGRVQFRFDTGSGPAVLTSLVPVEPGRWHRLELSRHWRQGTLSVDGETPVVGESPSGTDG 1377

Qy         145 VNAKGSVYIGGAPD---VATLTGGRFSSGITGCVKNLVLHSARPGAPPPQPLDLQHRAQA 201
               :|   ::|:|| |:      |       |: ||:: | :::  :     |   :|
Db        1378 LNLDTNLYVGGIPEEQVAMVLDRTSVGVGLKGCIRMLDINNQQLELSDWQRAAVQSSGVG 1437

Qy         202 GANTRPC 208
                    ||
Db        1438 ECGDHPC 1444

RESULT 12
US-08-460-309-2
; Sequence 2, Application US/08460309
; Patent No. 5837496
;  GENERAL INFORMATION:
;    APPLICANT:  Engvall, Eva
;    APPLICANT:  Leivo, Ilmo
;    TITLE OF INVENTION:  Nucleic Acids Encoding Merosin, Merosin
;    TITLE OF INVENTION:  Fragments and Uses Thereof
;    NUMBER OF SEQUENCES:  23
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Campbell and Flores
;      STREET:  4370 La Jolla Village Drive, Suite 700
;      CITY:  San Diego
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  92122
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/460,309
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/125,077
;      FILING DATE:  22-SEP-1993
;      APPLICATION NUMBER:  US PCT/US 94/10730
;      FILING DATE:  21-SEP-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/472,319
;      FILING DATE:  30-JAN-1990
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/919,951
;      FILING DATE:  27-JUL-1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Campbell, Cathryn A.
;      REGISTRATION NUMBER:  31,815
;      REFERENCE/DOCKET NUMBER:  P-LA 9721
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (619) 535-9001
;      TELEFAX:  (619) 535-8949
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1130 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-460-309-2

  Query Match             20.0%;  Score 223.5;  DB 2;  Length 1130;
  Best Local Similarity   36.4%;  Pred. No. 1e-13;
  Matches   59;  Conservative   22;  Mismatches   68;  Indels   13;  Gaps    5;

Qy          48 TIELEVRTSTASGLLLWQGVEVGEAGQGKDFISLGLQDGHLVFRYQLGSGEARLVSEDPI 107
               ||||||||   |||| :       |    || :: |::|   | | ||||:   :    |
Db         808 TIELEVRTEAESGLLFYMA-----AINHADFATVQLRNGLPYFSYDLGSGDTHTMIPTKI 862

Qy         108 NDGEWHRVTALREGRRGSIQVDGEELVSGRSPGPNVA--VNAKGSVYIGGAP-DVATLTG 164
               |||:||::  :|  : | : |||    | |:  |  |  ::  | :|:|| | :  |
Db         863 NDGQWHKIKIMRSKQEGILYVDG---ASNRTISPKKADILDVVGMLYVGGLPINYTTRRI 919

Qy         165 GRFSSGITGCVKNLVLHSARPGAPPPQPLDLQHRAQAGANTR 206
               | :   | ||||:|  || |   |   ||    |     || :
Db         920 GPVTYSIDGCVRN--LHMAEAPADLEQPTSSFHVGTCFANAQ 959

RESULT 13
US-08-125-077-2
; Sequence 2, Application US/08125077
; Patent No. 5872231
; Patent No. 5872231 5840863
;  GENERAL INFORMATION:
;    APPLICANT:  Engvall, Eva
;    APPLICANT:  Leivo, Ilmo
;    TITLE OF INVENTION:  Nucleic Acids Encoding Merosin, Merosin
;    TITLE OF INVENTION:  Fragments and Uses Thereof
;    NUMBER OF SEQUENCES:  23
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Campbell and Flores
;      STREET:  4370 La Jolla Village Drive, Suite 700
;      CITY:  San Diego
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  92122
;    COMPUTER READABLE FORM:
```

```
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.25
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/125,077
;       FILING DATE:  22-SEP-1993
;       CLASSIFICATION:  435
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US PCT/US 94/10730
;       FILING DATE:  21-SEP-1994
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 07/472,319
;       FILING DATE:  30-JAN-1990
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 07/919,951
;       FILING DATE:  27-JUL-1992
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Campbell, Cathryn A.
;       REGISTRATION NUMBER:  31,815
;       REFERENCE/DOCKET NUMBER:  P-LA 9721
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (619) 535-9001
;       TELEFAX:  (619) 535-8949
;   INFORMATION FOR SEQ ID NO:  2:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  1130 amino acids
;       TYPE:  amino acid
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  protein
US-08-125-077-2

  Query Match             20.0%;  Score 223.5;  DB 2;  Length 1130;
  Best Local Similarity   36.4%;  Pred. No. 1e-13;
  Matches   59;  Conservative   22;  Mismatches   68;  Indels   13;  Gaps    5;

Qy         48 TIELEVRTSTASGLLLWQGVEVGEAGQGKDFISLGLQDGHLVFRYQLGSGEARLVSEDPI 107
              ||||||||   |||| :       |    || :: |::|   | | ||||:   :    |
Db        808 TIELEVRTEAESGLLFYMA-----AINHADFATVQLRNGLPYFSYDLGSGDTHTMIPTKI 862

Qy        108 NDGEWHRVTALREGRRGSIQVDGEELVSGRSPGPNVA--VNAKGSVYIGGAP-DVATLTG 164
              |||:||::  :|  : | : |||    | |:  |  |  ::  | :|:|| | :  |
Db        863 NDGQWHKIKIMRSKQEGILYVDG---ASNRTISPKKADILDVVGMLYVGGLPINYTTRRI 919

Qy        165 GRFSSGITGCVKNLVLHSARPGAPPPQPLDLQHRAQAGANTR 206
              |  :  | |||:|  || |   |   ||    |     || :
Db        920 GPVTYSIDGCVRN--LHMAEAPADLEQPTSSFHVGTCFANAQ 959

RESULT 14
5444158-2
;Patent No. 5444158
;    APPLICANT: ENGVALL, EVA;SANES, JOSHUA
;    TITLE OF INVENTION: MEROSIN, NUCLEIC ACIDS ENCODING,
;FRAGMENTS AND USES THEREOF
;    NUMBER OF SEQUENCES: 4
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/87,642
;      FILING DATE: 08-JUL-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER: 587,689
;      FILING DATE: 24-SEP-1990
;      APPLICATION NUMBER: 472,319
;      FILING DATE: 30-JAN-1990
;SEQ ID NO:2:
;      LENGTH: 1130
5444158-2

  Query Match             20.0%;  Score 223.5;  DB 6;  Length 1130;
  Best Local Similarity   36.4%;  Pred. No. 1e-13;
  Matches   59;  Conservative   23;  Mismatches   67;  Indels   13;  Gaps    6;

Qy         48 TIELEVRTSTASGLLLWQGVEVGEAGQGKDFISLGLQDGHLVFRYQLGSGEARLVSEDPI 107
              ||||||||   |||| :    :  |    || :: |::|   | | ||||:   :    |
Db        808 TIELEVRTEAESGLLFYM-ARINHA----DFATVQLRNGLPYFSYDLGSGDTHTMIPTKI 862

Qy        108 NDGEWHRVTALREGRRGSIQVDGEELVSGRSPGPNVA--VNAKGSVYIGGAP-DVATLTG 164
              |||:||::  :|  : | : |||    | |:  |  |  ::  | :|:|| | :  |
Db        863 NDGQWHKIKIMRSKQEGILYVDG---ASNRTISPKKADILDVVGMLYVGGLPINYTTRRI 919

Qy        165 GRFSSGITGCVKNLVLHSARPGAPPPQPLDLQHRAQAGANTR 206
              |  :  | |||:|  || |   |   ||    |     || :
Db        920 GPVTYSIDGCVRN--LHMAEAPADLEQPTSSFHVGTCFANAQ 959

RESULT 15
US-09-562-702A-8
; Sequence 8, Application US/09562702A
; Patent No. 6632790
; GENERAL INFORMATION:
;  APPLICANT: Yurchenco, Peter
;  TITLE OF INVENTION: Laminin 2 and Methods for Its Use
;  FILE REFERENCE: 99-274-B
;  CURRENT APPLICATION NUMBER: US/09/562,702A
;  CURRENT FILING DATE:  2000-04-28
;  PRIOR APPLICATION NUMBER: 60/155,945
;  PRIOR FILING DATE: 1999-09-24
;  PRIOR APPLICATION NUMBER: 60/143,289
;  PRIOR FILING DATE: 1999-07-12
;  PRIOR APPLICATION NUMBER: 60/139,198
;  PRIOR FILING DATE: 1999-06-15
;  PRIOR APPLICATION NUMBER: 60/131,720
;  PRIOR FILING DATE: 1999-04-30
;  NUMBER OF SEQ ID NOS: 32
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 8
;   LENGTH: 3088
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-562-702A-8

  Query Match             20.0%;  Score 223.5;  DB 4;  Length 3088;
  Best Local Similarity   36.4%;  Pred. No. 4.3e-13;
  Matches   59;  Conservative   22;  Mismatches   68;  Indels   13;  Gaps    5;

Qy         48 TIELEVRTSTASGLLLWQGVEVGEAGQGKDFISLGLQDGHLVFRYQLGSGEARLVSEDPI 107
              ||||||||   |||| :       |    || :: |::|   | | ||||:   :    |
Db       2766 TIELEVRTEAESGLLFYMA-----AINHADFATVQLRNGLPYFSYDLGSGDTHTMIPTKI 2820

Qy        108 NDGEWHRVTALREGRRGSIQVDGEELVSGRSPGPNVA--VNAKGSVYIGGAP-DVATLTG 164
              |||:||::  :|  : | : |||    | |:  |  |  ::  | :|:|| | :  |
Db       2821 NDGQWHKIKIMRSKQEGILYVDG---ASNRTISPKKADILDVVGMLYVGGLPINYTTRRI 2877

Qy        165 GRFSSGITGCVKNLVLHSARPGAPPPQPLDLQHRAQAGANTR 206
              |  :  | |||:|  || |   |   ||    |     || :
Db       2878 GPVTYSIDGCVRN--LHMAEAPADLEQPTSSFHVGTCFANAQ 2917

Search completed: March  9, 2004, 17:23:34
Job time : 9.23186 secs
```

```
                    GenCore version 5.1.6
           Copyright (c) 1993 - 2004  Compugen Ltd.

OM protein - protein search, using sw model

Run on:          March  9, 2004, 17:19:08 ; Search time 15.4128 Seconds
                                            (without alignments)
                                            2876.963 Million cell updates/sec

Title:           US-10-006-011A-10
Perfect score:   1115
Sequence:        1 GIAESDWHLEGSGGNDAPGQ..........QPLDLQHRAQAGANTRPCPS 210

Scoring table:   BLOSUM62
                 Gapop 10.0 , Gapext 0.5

Searched:        809742 seqs, 211153259 residues

Total number of hits satisfying chosen parameters:      809742

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :       Published_Applications_AA:*
                1:  /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep:*
                2:  /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep:*
                3:  /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep:*
                4:  /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep:*
                5:  /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep:*
                6:  /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep:*
                7:  /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep:*
                8:  /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep:*
                9:  /cgn2_6/ptodata/1/pubpaa/US09A_PUBCOMB.pep:*
                10:  /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pep:*
                11:  /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep:*
                12:  /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep:*
                13:  /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep:*
                14:  /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep:*
                15:  /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep:*
                16:  /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep:*
                17:  /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep:*
                18:  /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep:*

      Pred. No. is the number of results predicted by chance to have a
      score greater than or equal to the score of the result being printed,
      and is derived by analysis of the total score distribution.

                               SUMMARIES
                 %
Result          Query
  No.   Score   Match Length DB  ID                            Description
-------------------------------------------------------------------------------
    1   278.5   25.0    152  15  US-10-104-047-3338             Sequence 3338, Ap
    2   278.5   25.0    210   9  US-09-764-870-432             Sequence 432, App
    3   278.5   25.0    210  14  US-10-125-540-432              Sequence 432, App
    4   278.5   25.0    238   9  US-09-764-853-758             Sequence 758, App
    5   278.5   25.0    238   9  US-09-764-898-252             Sequence 252, App
    6   278.5   25.0    238  10  US-09-764-881-158              Sequence 158, App
    7   278.5   25.0    238  14  US-10-073-865-121              Sequence 121, App
    8   278.5   25.0    238  15  US-10-242-747-158              Sequence 158, App
    9   278.5   25.0    375   9  US-09-978-249-10              Sequence 10, Appl
   10   278.5   25.0    432   9  US-09-764-853-541             Sequence 541, App
   11   278.5   25.0    432   9  US-09-764-898-179             Sequence 179, App
   12   278.5   25.0    432  10  US-09-764-881-102              Sequence 102, App
   13   278.5   25.0    432  14  US-10-073-865-78               Sequence 78, Appl
   14   278.5   25.0    432  15  US-10-242-747-102              Sequence 102, App
   15   278.5   25.0    463  15  US-10-104-047-3058             Sequence 3058, Ap
```

```
   16   278.5   25.0    775  15  US-10-108-260A-4433            Sequence 4433, Ap
   17   275.5   24.7    819  10  US-09-866-050A-503             Sequence 503, App
   18     265   23.8    406  15  US-10-264-049-2933             Sequence 2933, Ap
   19     265   23.8   1931  15  US-10-094-886-180              Sequence 180, App
   20     261   23.4    216  13  US-10-016-283-31               Sequence 31, Appl
   21     261   23.4    256  13  US-10-016-283-30               Sequence 30, Appl
   22     261   23.4    294  13  US-10-016-283-29               Sequence 29, Appl
   23     261   23.4    338  13  US-10-016-283-28               Sequence 28, Appl
   24     261   23.4    390  13  US-10-016-283-27               Sequence 27, Appl
   25     261   23.4    440  13  US-10-016-283-26               Sequence 26, Appl
   26     261   23.4    456  13  US-10-016-283-25               Sequence 25, Appl
   27     261   23.4    492  13  US-10-016-283-36               Sequence 36, Appl
   28   242.5   21.7   1940  13  US-10-016-283-34               Sequence 34, Appl
   29     240   21.5    194   9  US-09-764-870-537             Sequence 537, App
   30     240   21.5    194  14  US-10-125-540-537              Sequence 537, App
   31   223.5   20.0   3070  10  US-09-961-403-7                Sequence 7, Appli
   32   209.5   18.8    161  16  US-10-038-854-84               Sequence 84, Appl
   33     191   17.1   1712  15  US-10-291-172-359              Sequence 359, App
   34   188.5   16.9   3075   9  US-09-938-275-5               Sequence 5, Appli
   35   185.5   16.6   3672  15  US-10-369-493-6146             Sequence 6146, Ap
   36   170.5   15.3   4307  15  US-10-369-493-5698             Sequence 5698, Ap
   37   170.5   15.3   4307  15  US-10-369-493-5699             Sequence 5699, Ap
   38   170.5   15.3   4307  15  US-10-369-493-5700             Sequence 5700, Ap
   39     168   15.1   1438  15  US-10-297-639-2                Sequence 2, Appli
   40     168   15.1   2823  15  US-10-369-493-5220             Sequence 5220, Ap
   41     168   15.1   2823  15  US-10-369-493-5221             Sequence 5221, Ap
   42   165.5   14.8   1816  14  US-10-299-058-2                Sequence 2, Appli
   43   165.5   14.8   1816  14  US-10-299-058-4                Sequence 4, Appli
   44   165.5   14.8   1816  15  US-10-372-683-4                Sequence 4, Appli
   45     163   14.6   1061  15  US-10-291-265-347              Sequence 347, App

                                 ALIGNMENTS

RESULT 1
US-10-104-047-3338
; Sequence 3338, Application US/10104047
; Publication No. US20030236392A1
; GENERAL INFORMATION:
;  APPLICANT: HELIX RESEARCH INSTITUTE
;  TITLE OF INVENTION: No. US20030236392A1el full length cDNA
;  FILE REFERENCE: H1-A0105
;  CURRENT APPLICATION NUMBER: US/10/104,047
;  CURRENT FILING DATE:  2002-03-25
;  PRIOR APPLICATION NUMBER:
;  PRIOR FILING DATE:
;  NUMBER OF SEQ ID NOS: 4096
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3338
;   LENGTH: 152
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-104-047-3338

  Query Match             25.0%;  Score 278.5;  DB 15;  Length 152;
  Best Local Similarity   37.3%;  Pred. No. 5e-19;
  Matches   59;  Conservative   25;  Mismatches   65;  Indels    9;  Gaps    2;

Qy          51 LEVRTSTASGLLLWQGVEVGEAGQGKDFISLGLQDGHLVFRYQLGSGEARLVSEDPINDG 110
               :  :|:   |||||:|          |||||||:|| ||| | |||| | ::     |||
Db           1 MRFKTTAKDGLLLWRG--DSPMRPNSDFISLGLRDGALVFSYNLGSGVASIMVNGSFNDG 58

Qy         111 EWHRVTALREGRRGSIQVDGEELVSGRSPGPNVAVNAKGSVYIGGAPDVATLTGGRFSSG 170
                |||| |:|:|: | | ||     :|:|||    :|  |::|:||  ::|  |  ::  |
Db          59 RWHRVKAVRDGQSGKITVDDYGARTGKSPGMMRQLNINGALYVGGMKEIALHTNRQYMRG 118

Qy         171 ITGCVKNLVLHSARPGAPPPQPLDLQHRAQAGANTRPC 208
               : ||: :  | :          : |   |  | |   |
Db         119 LVGCISHFTLST-------DYHISLVEDAVDGKNINTC 149

RESULT 2
```

```
US-09-764-870-432
; Sequence 432, Application US/09764870
; Patent No. US20020042386A1
; GENERAL INFORMATION:
;  APPLICANT: Rosen et al.
;  TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
;  FILE REFERENCE: PTZ14
;  CURRENT APPLICATION NUMBER: US/09/764,870
;  CURRENT FILING DATE:  2001-01-17
;  Prior application data removed - consult PALM or file wrapper
;  NUMBER OF SEQ ID NOS: 646
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 432
;   LENGTH: 210
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-764-870-432

  Query Match             25.0%;  Score 278.5;  DB 9;  Length 210;
  Best Local Similarity   37.3%;  Pred. No. 7.6e-19;
  Matches   59;  Conservative   25;  Mismatches   65;  Indels    9;  Gaps    2;

Qy          51 LEVRTSTASGLLLWQGVEVGEAGQGKDFISLGLQDGHLVFRYQLGSGEARLVSEDPINDG 110
               :  :|:   |||||:|          |||||||:|| ||| | |||| | ::     |||
Db          59 MRFKTTAKDGLLLWRG--DSPMRPNSDFISLGLRDGALVFSYNLGSGVASIMVNGSFNDG 116

Qy         111 EWHRVTALREGRRGSIQVDGEELVSGRSPGPNVAVNAKGSVYIGGAPDVATLTGGRFSSG 170
                |||| |:|:|: | | ||     :|:|||    :|  |::|:||  ::|  |  ::  |
Db         117 RWHRVKAVRDGQSGKITVDDYGARTGKSPGMMRQLNINGALYVGGMKEIALHTNRQYMRG 176

Qy         171 ITGCVKNLVLHSARPGAPPPQPLDLQHRAQAGANTRPC 208
               : ||: :  | :          : |   |  | |   |
Db         177 LVGCISHFTLST-------DYHISLVEDAVDGKNINTC 207


RESULT 3
US-10-125-540-432
; Sequence 432, Application US/10125540
; Publication No. US20030059875A1
; GENERAL INFORMATION:
;  APPLICANT: Rosen et al.
;  TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
;  FILE REFERENCE: PTZ14C1
;  CURRENT APPLICATION NUMBER: US/10/125,540
;  CURRENT FILING DATE:  2002-04-19
;  Prior Application removed - See File Wrapper or Palm
;  NUMBER OF SEQ ID NOS: 646
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 432
;   LENGTH: 210
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-125-540-432

  Query Match             25.0%;  Score 278.5;  DB 14;  Length 210;
  Best Local Similarity   37.3%;  Pred. No. 7.6e-19;
  Matches   59;  Conservative   25;  Mismatches   65;  Indels    9;  Gaps    2;

Qy          51 LEVRTSTASGLLLWQGVEVGEAGQGKDFISLGLQDGHLVFRYQLGSGEARLVSEDPINDG 110
               :  :|:   |||||:|          |||||||:|| ||| | |||| | ::     |||
Db          59 MRFKTTAKDGLLLWRG--DSPMRPNSDFISLGLRDGALVFSYNLGSGVASIMVNGSFNDG 116

Qy         111 EWHRVTALREGRRGSIQVDGEELVSGRSPGPNVAVNAKGSVYIGGAPDVATLTGGRFSSG 170
                |||| |:|:|: | | ||     :|:|||    :|  |::|:||  ::|  |  ::  |
Db         117 RWHRVKAVRDGQSGKITVDDYGARTGKSPGMMRQLNINGALYVGGMKEIALHTNRQYMRG 176

Qy         171 ITGCVKNLVLHSARPGAPPPQPLDLQHRAQAGANTRPC 208
               : ||: :  | :          : |   |  | |   |
Db         177 LVGCISHFTLST-------DYHISLVEDAVDGKNINTC 207


RESULT 4
US-09-764-853-758
; Sequence 758, Application US/09764853
; Patent No. US20020090672A1
; GENERAL INFORMATION:
;  APPLICANT: Rosen et al.
;  TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
;  FILE REFERENCE: PJZ06
;  CURRENT APPLICATION NUMBER: US/09/764,853
;  CURRENT FILING DATE:  2001-01-17
;  Prior application data removed - consult PALM or file wrapper
;  NUMBER OF SEQ ID NOS: 939
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 758
;   LENGTH: 238
;   TYPE: PRT
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: SITE
;   LOCATION: (5)
;   OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-764-853-758

  Query Match             25.0%;  Score 278.5;  DB 9;  Length 238;
  Best Local Similarity   37.3%;  Pred. No. 9e-19;
  Matches   59;  Conservative   25;  Mismatches   65;  Indels    9;  Gaps    2;

Qy          51 LEVRTSTASGLLLWQGVEVGEAGQGKDFISLGLQDGHLVFRYQLGSGEARLVSEDPINDG 110
               :  :|:   |||||:|          |||||||:|| ||| | |||| | ::     |||
Db          87 MRFKTTAKDGLLLWRG--DSPMRPNSDFISLGLRDGALVFSYNLGSGVASIMVNGSFNDG 144

Qy         111 EWHRVTALREGRRGSIQVDGEELVSGRSPGPNVAVNAKGSVYIGGAPDVATLTGGRFSSG 170
                |||| |:|:|: | | ||     :|:|||    :|  |::|:||  ::|  |  ::  |
Db         145 RWHRVKAVRDGQSGKITVDDYGARTGKSPGMMRQLNINGALYVGGMKEIALHTNRQYMRG 204

Qy         171 ITGCVKNLVLHSARPGAPPPQPLDLQHRAQAGANTRPC 208
               : ||: :  | :          : |   |  | |   |
Db         205 LVGCISHFTLST-------DYHISLVEDAVDGKNINTC 235


RESULT 5
US-09-764-898-252
; Sequence 252, Application US/09764898
; Patent No. US20020090673A1
; GENERAL INFORMATION:
;  APPLICANT: Rosen et al.
;  TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
;  FILE REFERENCE: PJZ01
;  CURRENT APPLICATION NUMBER: US/09/764,898
;  CURRENT FILING DATE:  2001-01-17
;  Prior application data removed - consult PALM or file wrapper
;  NUMBER OF SEQ ID NOS: 311
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 252
;   LENGTH: 238
;   TYPE: PRT
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: SITE
;   LOCATION: (5)
;   OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-764-898-252

  Query Match             25.0%;  Score 278.5;  DB 9;  Length 238;
  Best Local Similarity   37.3%;  Pred. No. 9e-19;
  Matches   59;  Conservative   25;  Mismatches   65;  Indels    9;  Gaps    2;

Qy          51 LEVRTSTASGLLLWQGVEVGEAGQGKDFISLGLQDGHLVFRYQLGSGEARLVSEDPINDG 110
               :  :|:   |||||:|          |||||||:|| ||| | |||| | ::     |||
Db          87 MRFKTTAKDGLLLWRG--DSPMRPNSDFISLGLRDGALVFSYNLGSGVASIMVNGSFNDG 144

Qy         111 EWHRVTALREGRRGSIQVDGEELVSGRSPGPNVAVNAKGSVYIGGAPDVATLTGGRFSSG 170
                |||| |:|:|: | | ||     :|:|||    :|  |::|:||  ::|  |  ::  |
Db         145 RWHRVKAVRDGQSGKITVDDYGARTGKSPGMMRQLNINGALYVGGMKEIALHTNRQYMRG 204
```

```
Qy        171 ITGCVKNLVLHSARPGAPPPQPLDLQHRAQAGANTRPC 208
              : ||: :  | :          : |   |  | |   |
Db        205 LVGCISHFTLST-------DYHISLVEDAVDGKNINTC 235

RESULT 6
US-09-764-881-158
; Sequence 158, Application US/09764881
; Publication No. US20030125246A9
; GENERAL INFORMATION:
;  APPLICANT: Rosen et al.
;  TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
;  FILE REFERENCE: PTZ07
;  CURRENT APPLICATION NUMBER: US/09/764,881
;  CURRENT FILING DATE:  2001-01-17
;  Prior application data removed - refer to PALM or file wrapper
;  NUMBER OF SEQ ID NOS: 192
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 158
;   LENGTH: 238
;   TYPE: PRT
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: SITE
;   LOCATION: (5)
;   OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-764-881-158

  Query Match             25.0%;  Score 278.5;  DB 10;  Length 238;
  Best Local Similarity   37.3%;  Pred. No. 9e-19;
  Matches   59;  Conservative   25;  Mismatches   65;  Indels    9;  Gaps    2;

Qy         51 LEVRTSTASGLLLWQGVEVGEAGQGKDFISLGLQDGHLVFRYQLGSGEARLVSEDPINDG 110
              :  :|:   |||||:|          |||||||:|| ||| | |||| | ::     |||
Db         87 MRFKTTAKDGLLLWRG--DSPMRPNSDFISLGLRDGALVFSYNLGSGVASIMVNGSFNDG 144

Qy        111 EWHRVTALREGRRGSIQVDGEELVSGRSPGPNVAVNAKGSVYIGGAPDVATLTGGRFSSG 170
               |||| |:|:|: | | ||     :|:|||    :|  |::|:||  ::|  |   :: |
Db        145 RWHRVKAVRDGQSGKITVDDYGARTGKSPGMMRQLNINGALYVGGMKEIALHTNRQYMRG 204

Qy        171 ITGCVKNLVLHSARPGAPPPQPLDLQHRAQAGANTRPC 208
              : ||: :  | :          : |   |  | |   |
Db        205 LVGCISHFTLST-------DYHISLVEDAVDGKNINTC 235

RESULT 7
US-10-073-865-121
; Sequence 121, Application US/10073865
; Publication No. US20030044904A1
; GENERAL INFORMATION:
;  APPLICANT: Rosen et al.
;  TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
;  FILE REFERENCE: PJZ09C1
;  CURRENT APPLICATION NUMBER: US/10/073,865
;  CURRENT FILING DATE:  2002-02-14
;  Prior Application removed - See file Wrapper or Palm
;  NUMBER OF SEQ ID NOS: 154
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 121
;   LENGTH: 238
;   TYPE: PRT
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (5)
;   OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-10-073-865-121

  Query Match             25.0%;  Score 278.5;  DB 14;  Length 238;
  Best Local Similarity   37.3%;  Pred. No. 9e-19;
  Matches   59;  Conservative   25;  Mismatches   65;  Indels    9;  Gaps    2;

Qy         51 LEVRTSTASGLLLWQGVEVGEAGQGKDFISLGLQDGHLVFRYQLGSGEARLVSEDPINDG 110
              :  :|:   |||||:|          |||||||:|| ||| | |||| | ::     |||
Db         87 MRFKTTAKDGLLLWRG--DSPMRPNSDFISLGLRDGALVFSYNLGSGVASIMVNGSFNDG 144

Qy        111 EWHRVTALREGRRGSIQVDGEELVSGRSPGPNVAVNAKGSVYIGGAPDVATLTGGRFSSG 170
               |||| |:|:|: | | ||     :|:|||    :|  |::|:||  ::|  |   :: |
Db        145 RWHRVKAVRDGQSGKITVDDYGARTGKSPGMMRQLNINGALYVGGMKEIALHTNRQYMRG 204

Qy        171 ITGCVKNLVLHSARPGAPPPQPLDLQHRAQAGANTRPC 208
              : ||: :  | :          : |   |  | |   |
Db        205 LVGCISHFTLST-------DYHISLVEDAVDGKNINTC 235

RESULT 8
US-10-242-747-158
; Sequence 158, Application US/10242747
; Publication No. US20040005577A1
; GENERAL INFORMATION:
;  APPLICANT: Rosen et al.
;  TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
;  FILE'REFERENCE: PTZ07C1
;  CURRENT APPLICATION NUMBER: US/10/242,747
;  CURRENT FILING DATE:  2002-09-13
;  PRIOR APPLICATION NUMBER: 09/764,881
;  PRIOR FILING DATE: 2001-01-17
;  PRIOR APPLICATION NUMBER: 60/179,065
;  PRIOR FILING DATE: 2000-01-31
;  PRIOR APPLICATION NUMBER: 60/180,628
;  PRIOR FILING DATE: 2000-02-04
;  PRIOR APPLICATION NUMBER: 60/214,886
;  PRIOR FILING DATE: 2000-06-28
;  PRIOR APPLICATION NUMBER: 60/217,487
;  PRIOR FILING DATE: 2000-07-11
;  PRIOR APPLICATION NUMBER: 60/225,758
;  PRIOR FILING DATE: 2000-08-14
;  PRIOR APPLICATION NUMBER: 60/220,963
;  PRIOR FILING DATE: 2000-07-26
;  PRIOR APPLICATION NUMBER: 60/217,496
;  PRIOR FILING DATE: 2000-07-11
;  PRIOR APPLICATION NUMBER: 60/225,447
;  PRIOR FILING DATE: 2000-08-14
;  PRIOR APPLICATION NUMBER: 60/218,290
;  PRIOR FILING DATE: 2000-07-14
;  Remaining Prior Application data removed - See File Wrapper or PALM.
;  NUMBER OF SEQ ID NOS: 192
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 158
;   LENGTH: 238
;   TYPE: PRT
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (5)
;   OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-10-242-747-158

  Query Match             25.0%;  Score 278.5;  DB 15;  Length 238;
  Best Local Similarity   37.3%;  Pred. No. 9e-19;
  Matches   59;  Conservative   25;  Mismatches   65;  Indels    9;  Gaps    2;

Qy         51 LEVRTSTASGLLLWQGVEVGEAGQGKDFISLGLQDGHLVFRYQLGSGEARLVSEDPINDG 110
              :  :|:   |||||:|          |||||||:|| ||| | |||| | ::     |||
Db         87 MRFKTTAKDGLLLWRG--DSPMRPNSDFISLGLRDGALVFSYNLGSGVASIMVNGSFNDG 144

Qy        111 EWHRVTALREGRRGSIQVDGEELVSGRSPGPNVAVNAKGSVYIGGAPDVATLTGGRFSSG 170
               |||| |:|:|: | | ||     :|:|||    :|  |::|:||  ::|  |   :: |
Db        145 RWHRVKAVRDGQSGKITVDDYGARTGKSPGMMRQLNINGALYVGGMKEIALHTNRQYMRG 204

Qy        171 ITGCVKNLVLHSARPGAPPPQPLDLQHRAQAGANTRPC 208
              : ||: :  | :          : |   |  | |   |
Db        205 LVGCISHFTLST-------DYHISLVEDAVDGKNINTC 235

RESULT 9
```

```
US-09-978-249-10
; Sequence 10, Application US/09978249
; Patent No. US20020106780A1
; GENERAL INFORMATION:
;  APPLICANT: Fiscella, et al.
;  TITLE OF INVENTION: Extracellular Matrix Polynucleotides, Polypeptides, and Antibodie
;  FILE REFERENCE: PT054P1
;  CURRENT APPLICATION NUMBER: US/09/978,249
;  CURRENT FILING DATE:  2001-10-17
;  PRIOR APPLICATION NUMBER: PCT/US01/11643
;  PRIOR FILING DATE: 2001-04-11
;  PRIOR APPLICATION NUMBER: 60/198,123
;  PRIOR FILING DATE: 2000-04-18
;  NUMBER OF SEQ ID NOS: 16
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 10
;   LENGTH: 375
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-978-249-10

  Query Match              25.0%;  Score 278.5;  DB 9;  Length 375;
  Best Local Similarity   37.3%;  Pred. No. 1.6e-18;
  Matches   59;  Conservative   25;  Mismatches   65;  Indels    9;  Gaps    2;

Qy          51 LEVRTSTASGLLLWQGVEVGEAGQGKDFISLGLQDGHLVFRYQLGSGEARLVSEDPINDG 110
               :  :|:   |||||:|          |||||||:|| ||| | |||| | ::     |||
Db         224 MRFKTTAKDGLLLWRG--DSPMRPNSDFISLGLRDGALVFSYNLGSGVASIMVNGSFNDG 281

Qy         111 EWHRVTALREGRRGSIQVDGEELVSGRSPGPNVAVNAKGSVYIGGAPDVATLTGGRFSSG 170
                |||| |:|:|: | | ||     :|:|||    :|  |::|:||  ::|  |  ::  |
Db         282 RWHRVKAVRDGQSGKITVDDYGARTGKSPGMMRQLNINGALYVGGMKEIALHTNRQYMRG 341

Qy         171 ITGCVKNLVLHSARPGAPPPQPLDLQHRAQAGANTRPC 208
               : ||: :  | :          : |   |  | |   |
Db         342 LVGCISHFTLST-------DYHISLVEDAVDGKNINTC 372


RESULT 10
US-09-764-853-541
; Sequence 541, Application US/09764853
; Patent No. US20020090672A1
; GENERAL INFORMATION:
;  APPLICANT: Rosen et al.
;  TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
;  FILE REFERENCE: PJZ06
;  CURRENT APPLICATION NUMBER: US/09/764,853
;  CURRENT FILING DATE:  2001-01-17
;  Prior application data removed - consult PALM or file wrapper
;  NUMBER OF SEQ ID NOS: 939
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 541
;   LENGTH: 432
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-764-853-541

  Query Match              25.0%;  Score 278.5;  DB 9;  Length 432;
  Best Local Similarity   37.3%;  Pred. No. 2e-18;
  Matches   59;  Conservative   25;  Mismatches   65;  Indels    9;  Gaps    2;

Qy          51 LEVRTSTASGLLLWQGVEVGEAGQGKDFISLGLQDGHLVFRYQLGSGEARLVSEDPINDG 110
               :  :|:   |||||:|          |||||||:|| ||| | |||| | ::     |||
Db         281 MRFKTTAKDGLLLWRG--DSPMRPNSDFISLGLRDGALVFSYNLGSGVASIMVNGSFNDG 338

Qy         111 EWHRVTALREGRRGSIQVDGEELVSGRSPGPNVAVNAKGSVYIGGAPDVATLTGGRFSSG 170
                |||| |:|:|: | | ||     :|:|||    :|  |::|:||  ::|  |  ::  |
Db         339 RWHRVKAVRDGQSGKITVDDYGARTGKSPGMMRQLNINGALYVGGMKEIALHTNRQYMRG 398

Qy         171 ITGCVKNLVLHSARPGAPPPQPLDLQHRAQAGANTRPC 208
               : ||: :  | :          : |   |  | |   |
Db         399 LVGCISHFTLST-------DYHISLVEDAVDGKNINTC 429


RESULT 11
US-09-764-898-179
; Sequence 179, Application US/09764898
; Patent No. US20020090673A1
; GENERAL INFORMATION:
;  APPLICANT: Rosen et al.
;  TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
;  FILE REFERENCE: PJZ01
;  CURRENT APPLICATION NUMBER: US/09/764,898
;  CURRENT FILING DATE:  2001-01-17
;  Prior application data removed - consult PALM or file wrapper
;  NUMBER OF SEQ ID NOS: 311
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 179
;   LENGTH: 432
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-764-898-179

  Query Match              25.0%;  Score 278.5;  DB 9;  Length 432;
  Best Local Similarity   37.3%;  Pred. No. 2e-18;
  Matches   59;  Conservative   25;  Mismatches   65;  Indels    9;  Gaps    2;

Qy          51 LEVRTSTASGLLLWQGVEVGEAGQGKDFISLGLQDGHLVFRYQLGSGEARLVSEDPINDG 110
               :  :|:   |||||:|          |||||||:|| ||| | |||| | ::     |||
Db         281 MRFKTTAKDGLLLWRG--DSPMRPNSDFISLGLRDGALVFSYNLGSGVASIMVNGSFNDG 338

Qy         111 EWHRVTALREGRRGSIQVDGEELVSGRSPGPNVAVNAKGSVYIGGAPDVATLTGGRFSSG 170
                |||| |:|:|: | | ||     :|:|||    :|  |::|:||  ::|  |  ::  |
Db         339 RWHRVKAVRDGQSGKITVDDYGARTGKSPGMMRQLNINGALYVGGMKEIALHTNRQYMRG 398

Qy         171 ITGCVKNLVLHSARPGAPPPQPLDLQHRAQAGANTRPC 208
               : ||: :  | :          : |   |  | |   |
Db         399 LVGCISHFTLST-------DYHISLVEDAVDGKNINTC 429


RESULT 12
US-09-764-881-102
; Sequence 102, Application US/09764881
; Publication No. US20030125246A9
; GENERAL INFORMATION:
;  APPLICANT: Rosen et al.
;  TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
;  FILE REFERENCE: PTZ07
;  CURRENT APPLICATION NUMBER: US/09/764,881
;  CURRENT FILING DATE:  2001-01-17
;  Prior application data removed - refer to PALM or file wrapper
;  NUMBER OF SEQ ID NOS: 192
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 102
;   LENGTH: 432
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-764-881-102

  Query Match              25.0%;  Score 278.5;  DB 10;  Length 432;
  Best Local Similarity   37.3%;  Pred. No. 2e-18;
  Matches   59;  Conservative   25;  Mismatches   65;  Indels    9;  Gaps    2;

Qy          51 LEVRTSTASGLLLWQGVEVGEAGQGKDFISLGLQDGHLVFRYQLGSGEARLVSEDPINDG 110
               :  :|:   |||||:|          |||||||:|| ||| | |||| | ::     |||
Db         281 MRFKTTAKDGLLLWRG--DSPMRPNSDFISLGLRDGALVFSYNLGSGVASIMVNGSFNDG 338

Qy         111 EWHRVTALREGRRGSIQVDGEELVSGRSPGPNVAVNAKGSVYIGGAPDVATLTGGRFSSG 170
                |||| |:|:|: | | ||     :|:|||    :|  |::|:||  ::|  |  ::  |
Db         339 RWHRVKAVRDGQSGKITVDDYGARTGKSPGMMRQLNINGALYVGGMKEIALHTNRQYMRG 398

Qy         171 ITGCVKNLVLHSARPGAPPPQPLDLQHRAQAGANTRPC 208
               : ||: :  | :          : |   |  | |   |
Db         399 LVGCISHFTLST-------DYHISLVEDAVDGKNINTC 429
```

```
RESULT 13
US-10-073-865-78
; Sequence 78, Application US/10073865
; Publication No. US20030044904A1
; GENERAL INFORMATION:
;  APPLICANT: Rosen et al.
;  TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
;  FILE REFERENCE: PJZ09C1
;  CURRENT APPLICATION NUMBER: US/10/073,865
;  CURRENT FILING DATE:  2002-02-14
;  Prior Application removed - See file Wrapper or Palm
;  NUMBER OF SEQ ID NOS: 154
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 78
;   LENGTH: 432
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-073-865-78

  Query Match             25.0%;  Score 278.5;  DB 14;  Length 432;
  Best Local Similarity   37.3%;  Pred. No. 2e-18;
  Matches   59;  Conservative   25;  Mismatches   65;  Indels    9;  Gaps    2;

Qy          51 LEVRTSTASGLLLWQGVEVGEAGQGKDFISLGLQDGHLVFRYQLGSGEARLVSEDPINDG 110
               :  :|:   |||||:|          |||||||:|| ||| | |||| | ::     |||
Db         281 MRFKTTAKDGLLLWRG--DSPMRPNSDFISLGLRDGALVFSYNLGSGVASIMVNGSFNDG 338

Qy         111 EWHRVTALREGRRGSIQVDGEELVSGRSPGPNVAVNAKGSVYIGGAPDVATLTGGRFSSG 170
                |||| |:|:|: | | ||     :|:|||    :|  |::|:||  ::|  |  ::  |
Db         339 RWHRVKAVRDGQSGKITVDDYGARTGKSPGMMRQLNINGALYVGGMKEIALHTNRQYMRG 398

Qy         171 ITGCVKNLVLHSARPGAPPPQPLDLQHRAQAGANTRPC 208
               : ||: :  | :          : |   |  | |   |
Db         399 LVGCISHFTLST-------DYHISLVEDAVDGKNINTC 429

RESULT 14
US-10-242-747-102
; Sequence 102, Application US/10242747
; Publication No. US20040005577A1
; GENERAL INFORMATION:
;  APPLICANT: Rosen et al.
;  TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
;  FILE REFERENCE: PTZ07C1
;  CURRENT APPLICATION NUMBER: US/10/242,747
;  CURRENT FILING DATE:  2002-09-13
;  PRIOR APPLICATION NUMBER: 09/764,881
;  PRIOR FILING DATE: 2001-01-17
;  PRIOR APPLICATION NUMBER: 60/179,065
;  PRIOR FILING DATE: 2000-01-31
;  PRIOR APPLICATION NUMBER: 60/180,628
;  PRIOR FILING DATE: 2000-02-04
;  PRIOR APPLICATION NUMBER: 60/214,886
;  PRIOR FILING DATE: 2000-06-28
;  PRIOR APPLICATION NUMBER: 60/217,487
;  PRIOR FILING DATE: 2000-07-11
;  PRIOR APPLICATION NUMBER: 60/225,758
;  PRIOR FILING DATE: 2000-08-14
;  PRIOR APPLICATION NUMBER: 60/220,963
;  PRIOR FILING DATE: 2000-07-26
;  PRIOR APPLICATION NUMBER: 60/217,496
;  PRIOR FILING DATE: 2000-07-11
;  PRIOR APPLICATION NUMBER: 60/225,447
;  PRIOR FILING DATE: 2000-08-14
;  PRIOR APPLICATION NUMBER: 60/218,290
;  PRIOR FILING DATE: 2000-07-14
;  Remaining Prior Application data removed - See File Wrapper or PALM.
;  NUMBER OF SEQ ID NOS: 192
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 102
;   LENGTH: 432
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-242-747-102

  Query Match             25.0%;  Score 278.5;  DB 15;  Length 432;
  Best Local Similarity   37.3%;  Pred. No. 2e-18;
  Matches   59;  Conservative   25;  Mismatches   65;  Indels    9;  Gaps    2;

Qy          51 LEVRTSTASGLLLWQGVEVGEAGQGKDFISLGLQDGHLVFRYQLGSGEARLVSEDPINDG 110
               :  :|:   |||||:|          |||||||:|| ||| | |||| | ::     |||
Db         281 MRFKTTAKDGLLLWRG--DSPMRPNSDFISLGLRDGALVFSYNLGSGVASIMVNGSFNDG 338

Qy         111 EWHRVTALREGRRGSIQVDGEELVSGRSPGPNVAVNAKGSVYIGGAPDVATLTGGRFSSG 170
                |||| |:|:|: | | ||     :|:|||    :|  |::|:||  ::|  |  ::  |
Db         339 RWHRVKAVRDGQSGKITVDDYGARTGKSPGMMRQLNINGALYVGGMKEIALHTNRQYMRG 398

Qy         171 ITGCVKNLVLHSARPGAPPPQPLDLQHRAQAGANTRPC 208
               : ||: :  | :          : |   |  | |   |
Db         399 LVGCISHFTLST-------DYHISLVEDAVDGKNINTC 429

RESULT 15
US-10-104-047-3058
; Sequence 3058, Application US/10104047
; Publication No. US20030236392A1
; GENERAL INFORMATION:
;  APPLICANT: HELIX RESEARCH INSTITUTE
;  TITLE OF INVENTION: No. US20030236392A1el full length cDNA
;  FILE REFERENCE: H1-A0105
;  CURRENT APPLICATION NUMBER: US/10/104,047
;  CURRENT FILING DATE:  2002-03-25
;  PRIOR APPLICATION NUMBER:
;  PRIOR FILING DATE:
;  NUMBER OF SEQ ID NOS: 4096
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3058
;   LENGTH: 463
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-104-047-3058

  Query Match             25.0%;  Score 278.5;  DB 15;  Length 463;
  Best Local Similarity   37.3%;  Pred. No. 2.1e-18;
  Matches   59;  Conservative   25;  Mismatches   65;  Indels    9;  Gaps    2;

Qy          51 LEVRTSTASGLLLWQGVEVGEAGQGKDFISLGLQDGHLVFRYQLGSGEARLVSEDPINDG 110
               :  :|:   |||||:|          |||||||:|| ||| | |||| | ::     |||
Db         312 MRFKTTAKDGLLLWRG--DSPMRPNSDFISLGLRDGALVFSYNLGSGVASIMVNGSFNDG 369

Qy         111 EWHRVTALREGRRGSIQVDGEELVSGRSPGPNVAVNAKGSVYIGGAPDVATLTGGRFSSG 170
                |||| |:|:|: | | ||     :|:|||    :|  |::|:||  ::|  |  ::  |
Db         370 RWHRVKAVRDGQSGKITVDDYGARTGKSPGMMRQLNINGALYVGGMKEIALHTNRQYMRG 429

Qy         171 ITGCVKNLVLHSARPGAPPPQPLDLQHRAQAGANTRPC 208
               : ||: :  | :          : |   |  | |   |
Db         430 LVGCISHFTLST-------DYHISLVEDAVDGKNINTC 460

Search completed: March  9, 2004, 17:25:16
Job time : 16.4128 secs
```

```
                   GenCore version 5.1.6
          Copyright (c) 1993 - 2004  Compugen Ltd.

OM protein - protein search, using sw model

Run on:          March  9, 2004, 17:15:33 ; Search time 7.18098 Seconds
                                           (without alignments)
                                           2813.016 Million cell updates/sec

Title:           US-10-006-011A-10
Perfect score:   1115
Sequence:        1 GIAESDWHLEGSGGNDAPGQ..........QPLDLQHRAQAGANTRPCPS 210

Scoring table:   BLOSUM62
                 Gapop 10.0 , Gapext 0.5

Searched:        283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters:      283366

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :        PIR_78:*
                 1:  pir1:*
                 2:  pir2:*
                 3:  pir3:*
                 4:  pir4:*

      Pred. No. is the number of results predicted by chance to have a
      score greater than or equal to the score of the result being printed,
      and is derived by analysis of the total score distribution.

                                 SUMMARIES
                  %
Result          Query
  No.   Score   Match Length DB  ID                    Description
------------------------------------------------------------------------------
    1    1115   100.0   4391  2  A38096                perlecan precursor
    2   964.5    86.5   3707  2  S18252                heparan sulfate pr
    3   338.5    30.4   3375  2  T19821                hypothetical prote
    4   242.5    21.7   1959  1  AGRT                  agrin - rat
    5     235    21.1   1328  2  T43060                agrin - electric r
    6   223.5    20.0   1751  1  MMHUMH                laminin alpha-2 ch
    7   207.5    18.6   3106  1  S53868                laminin alpha-2 ch
    8   200.5    18.0   1955  1  AGCH                  agrin precursor -
    9   188.5    16.9   3075  2  S14458                laminin alpha-1 ch
   10     188    16.9   1715  2  C40228                neurexin II-alpha
   11   185.5    16.6   3672  2  T23433                hypothetical prote
   12   185.5    16.6   3704  2  T37316                probable laminin a
   13   170.5    15.3   4307  2  T20721                hypothetical prote
   14     168    15.1   1438  2  A48216                neurexin III-alpha
   15     168    15.1   1471  2  B48218                neurexin III-alpha
   16     168    15.1   1578  2  I48216                neurexin III-alpha
   17     168    15.1   2823  2  T23064                hypothetical prote
   18     168    15.1   2823  2  F87908                protein T22A3.8 [i
   19     168    15.1   3102  2  T43291                laminin alpha chai
   20     167    15.0   1264  2  T19545                hypothetical prote
   21   165.5    14.8   1816  1  S68960                laminin alpha-4 ch
   22   162.5    14.6   1713  2  A55347                adhesive ligand ep
   23     162    14.5   3084  1  MMMSA                 laminin alpha-1 ch
   24   160.5    14.4   3712  2  S18253                laminin alpha-1 ch
   25   159.5    14.3   1531  2  T42218                slit-1 protein hom
   26     159    14.3   1507  2  A40228                neurexin I-alpha p
   27     159    14.3   1530  2  I45944                neurexin I-alpha -
   28     153    13.7    223  2  T16335                hypothetical prote
   29   148.5    13.3   2809  2  T30213                G-cadherin - sea u
   30     147    13.2   1025  2  T42626                secreted leucine-r
   31   143.5    12.9   1647  2  T32934                hypothetical prote
   32   139.5    12.5    601  2  T22025                hypothetical prote
   33   139.5    12.5    601  2  D89711                protein F40E10.4 [
   34     139    12.5   1283  2  T13799                neurexin IV - frui
   35   134.5    12.1   5147  1  IJFFTM                cadherin-related t
   36   132.5    11.9    147  2  I54245                laminin M - rat (f
   37   130.5    11.7   1381  2  T31083                paranodin - rat
   38   123.5    11.1   1385  2  T14158                neurexin IV - mous
   39   123.5    11.1   3097  2  T00021                DN-cadherin - frui
   40     123    11.0   1216  2  T26104                hypothetical prote
   41   122.5    11.0   2610  2  T20968                hypothetical prote
   42     119    10.7   1523  2  T13953                MEGF5 protein - ra
   43   116.5    10.4   3635  2  T10053                laminin alpha 5 ch
   44   113.5    10.2    648  2  T34999                probable neuramida
   45     113    10.1    402  1  BOHUS                 sex steroid-bindin

                                 ALIGNMENTS

RESULT 1
A38096
perlecan precursor - human
N;Alternate names: basement membrane heparan sulfate proteoglycan; heparan sulfate prote
C;Species: Homo sapiens (man)
C;Date: 07-Apr-1994 #sequence_revision 07-Apr-1994 #text_change 05-Nov-1999
C;Accession: A38096; S19256; S77946; A41059; A40306; B33625; A33625; A41736
R;Murdoch, A.D.; Dodge, G.R.; Cohen, I.; Tuan, R.S.; Iozzo, R.V.
J. Biol. Chem. 267, 8544-8557, 1992
A;Title: Primary structure of the human heparan sulfate proteoglycan from basement membr
tor, laminin, neural cell adhesion molecules, and epidermal growth factor.
A;Reference number: A38096; MUID:92235084; PMID:1569102
A;Accession: A38096
A;Molecule type: mRNA
A;Residues: 1-4391 <MUR>
A;Cross-references: GB:M85289; NID:g184426; PIDN:AAA52700.1; PID:g184427
R;Kallunki, P.; Tryggvason, K.
J. Cell Biol. 116, 559-571, 1992
A;Title: Human basement membrane heparan sulfate proteoglycan core protein: a 467-kD prc
ell adhesion molecules, and epidermal growth factor.
A;Reference number: A41736; MUID:92112994; PMID:1730768
A;Accession: S19256
A;Molecule type: mRNA
A;Residues: 1-57,'D',59-434,'A',436,'FL',438-449,'Q',451-502,'A',503-792,'K',794-908,'R'
71-2979,'H',2981-2994,'G',2996-3167,'T',3169-3240,'R',3242-3426,'R',3428-3631,'Q',3633-3
A;Cross-references: EMBL:X62515
R;Tryggvason, K.
submitted to the EMBL Data Library, October 1991
A;Reference number: S77946
A;Accession: S77946
A;Molecule type: mRNA
A;Residues: 1-57,'D',59-434,'A',436,'FL',438-449,'Q',451-502,'A',503-792,'K',794-908,'R'
71-2979,'H',2981-2994,'G',2996-3167,'T',3169-3240,'R',3242-3426,'R',3428-3631,'Q',3633-4
A;Cross-references: EMBL:X62515; NID:g29469; PIDN:CAA44373.1; PID:g29470
R;Kallunki, P.; Eddy, R.L.; Byers, M.G.; Kestilae, M.; Shows, T.B.; Tryggvason, K.
Genomics 11, 389-396, 1991
A;Title: Cloning of human heparan sulfate proteoglycan core protein, assignment of the g
A;Reference number: A41059; MUID:92120660; PMID:1685141
A;Accession: A41059
A;Molecule type: mRNA
A;Residues: 'RT',892-908,'R',910-1101,'L',1103-1132,'L',1134-1221,'L',1223-1397 <KA2>
A;Cross-references: GB:S76436; NID:g243370; PIDN:AAB21121.1; PID:g243371
R;Dodge, G.R.; Kovalszky, I.; Chu, M.L.; Hassell, J.R.; McBride, O.W.; Yi, H.F.; Iozzo,
Genomics 10, 673-680, 1991
A;Title: Heparan sulfate proteoglycan of human colon: partial molecular cloning, cellula
A;Reference number: A40306; MUID:91365376; PMID:1679749
A;Accession: A40306
A;Molecule type: mRNA
A;Residues: 1018-1405,'G',1407-1409,'G',1411-1465 <DOD>
A;Cross-references: GB:M64283; NID:g184424; PIDN:AAA52699.1; PID:g184425
R;Heremans, A.; van der Schueren, B.; De Cock, B.; Paulsson, M.; Cassiman, J.J.; van den
J. Cell Biol. 109, 3199-3211, 1989
```

```
A;Title: Matrix-associated heparan sulfate proteoglycan: core protein-specific monoclona
anes.
A;Reference number: A33625; MUID:90078352; PMID:2687294
A;Accession: B33625
A;Molecule type: protein
A;Residues: 1379-1384,'X',1386-1388,'X',1390-1398 <HE2>
A;Accession: A33625
A;Molecule type: protein
A;Residues: 2166-2171,'X',2173-2175,'X',2177-2185 <HE3>
A;Note: peptide potentially matches four different regions of sequence shown
C;Genetics:
A;Gene: GDB:HSPG2
A;Cross-references: GDB:126372; OMIM:142461
A;Map position: 1p36.1-1p36.1
C;Superfamily: LDL receptor ligand-binding repeat homology; EGF homology; laminin G repe
C;Keywords: chondroitin sulfate proteoglycan; glycoprotein; heparan sulfate; transmembra
F;1-21/Domain: signal sequence #status predicted <SIG>
F;22-4391/Product: perlecan #status predicted <MAT>
F;22-193/Domain: I <DOM1>
F;194-530/Domain: II <DOM2>
F;199-234/Domain: LDL receptor ligand-binding repeat homology <LDL1>
F;285-319/Domain: LDL receptor ligand-binding repeat homology <LDL2>
F;325-359/Domain: LDL receptor ligand-binding repeat homology <LDL3>
F;368-403/Domain: LDL receptor ligand-binding repeat homology <LDL4>
F;531-1676/Domain: III <DOM3>
F;1159-1206/Domain: laminin-type EGF-like homology <LEG>
F;1563-1610/Domain: laminin-type EGF-like homology <EG7>
F;1613-1668/Domain: laminin-type EGF-like homology <LEG8>
F;1677-3686/Domain: IV <DOM4>
F;2007-2034/Domain: transmembrane #status predicted <TRM>
F;3687-4391/Domain: V <DOM5>
F;3845-3880/Domain: EGF homology <EGF1>
F;3888-3921/Domain: EGF homology <EGF>
F;3953-4106/Domain: laminin G repeat homology <LG2>
F;4147-4175/Domain: EGF homology <EGF2>
F;4149-4151/Region: motor neuron attachment (L-R-E) motif
F;4299-4301/Region: motor neuron attachment (L-R-E) motif
F;65,71,76/Binding site: heparan sulfate (Ser) (covalent) #status predicted
F;89,554,1755,2121,3072,3105,3279,3780,3836,4068/Binding site: carbohydrate (Asn) (coval
F;2995,3933,4179/Binding site: chondroitin sulfate (Ser) (covalent) #status predicted

  Query Match             100.0%;  Score 1115;  DB 2;  Length 4391;
  Best Local Similarity   100.0%;  Pred. No. 9.4e-83;
  Matches  210;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 GIAESDWHLEGSGGNDAPGQYGAYFHDDGFLAFPGHVFSRSLPEVPETIELEVRTSTASG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4182 GIAESDWHLEGSGGNDAPGQYGAYFHDDGFLAFPGHVFSRSLPEVPETIELEVRTSTASG 4241

Qy         61 LLLWQGVEVGEAGQGKDFISLGLQDGHLVFRYQLGSGEARLVSEDPINDGEWHRVTALRE 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4242 LLLWQGVEVGEAGQGKDFISLGLQDGHLVFRYQLGSGEARLVSEDPINDGEWHRVTALRE 4301

Qy        121 GRRGSIQVDGEELVSGRSPGPNVAVNAKGSVYIGGAPDVATLTGGRFSSGITGCVKNLVL 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4302 GRRGSIQVDGEELVSGRSPGPNVAVNAKGSVYIGGAPDVATLTGGRFSSGITGCVKNLVL 4361

Qy        181 HSARPGAPPPQPLDLQHRAQAGANTRPCPS 210
              ||||||||||||||||||||||||||||||
Db       4362 HSARPGAPPPQPLDLQHRAQAGANTRPCPS 4391

RESULT 2
S18252
heparan sulfate proteoglycan - mouse
N;Alternate names: perlecan
C;Species: Mus musculus (house mouse)
C;Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 05-Nov-1999
C;Accession: S18252; A31917; B31917; S66460
R;Noonan, D.M.; Fulle, A.; Valente, P.; Cai, S.; Horigan, E.; Sasaki, M.; Yamada, Y.; Ha
J. Biol. Chem. 266, 22939-22947, 1991
A;Title: The complete sequence of perlecan, a basement membrane heparan sulfate proteogl
adhesion molecule.
A;Reference number: S18252; MUID:92078153; PMID:1744087
A;Accession: S18252
A;Molecule type: mRNA
A;Residues: 1-3707 <NOO>
A;Cross-references: EMBL:M77174; NID:g200295; PIDN:AAA39911.1; PID:g200296
R;Noonan, D.M.; Horigan, E.A.; Ledbetter, S.R.; Vogeli, G.; Sasaki, M.; Yamada, Y.; Hass
J. Biol. Chem. 263, 16379-16387, 1988
A;Title: Identification of cDNA clones encoding different domains of the basement membra
A;Reference number: A92680; MUID:89034110; PMID:2972708
A;Accession: A31917
A;Molecule type: mRNA
A;Residues: 940-1601 <NO2>
A;Cross-references: GB:J04054; NID:g200252; PIDN:AAA39899.1; PID:g200253
A;Accession: B31917
A;Molecule type: mRNA
A;Residues: 1870-2600 <NO3>
A;Cross-references: GB:J04055; NID:g200300; PIDN:AAA39912.1; PID:g200301
R;Schulze, B.; Mann, K.; Battistutta, R.; Wiedemann, H.; Timpl, R.
Eur. J. Biochem. 231, 551-556, 1995
A;Title: Structural properties of recombinant domain III-3 of perlecan containing a glob
A;Reference number: S66460; MUID:95377282; PMID:7649154
A;Accession: S66460
A;Molecule type: protein
A;Residues: 1272-1274,'X',1276,'X',1278-1279 <SCH>
C;Superfamily: LDL receptor ligand-binding repeat homology; EGF homology; laminin G repe
C;Keywords: glycoprotein
F;199-234/Domain: LDL receptor ligand-binding repeat homology <LDL1>
F;285-319/Domain: LDL receptor ligand-binding repeat homology <LDL2>
F;325-359/Domain: LDL receptor ligand-binding repeat homology <LDL3>
F;368-403/Domain: LDL receptor ligand-binding repeat homology <LDL4>
F;764-811/Domain: laminin-type EGF-like homology <LEG>
F;1159-1206/Domain: laminin-type EGF-like homology <LEG7>
F;1563-1610/Domain: laminin-type EGF-like homology <EG7>
F;1613-1668/Domain: laminin-type EGF-like homology <LEG8>
F;3163-3198/Domain: EGF homology <EGF>
F;3270-3423/Domain: laminin G repeat homology <LG2>
F;3464-3492/Domain: EGF homology <EGF7>
F;1256,1891,2336,2394,2427/Binding site: carbohydrate (Asn) (covalent) #status predictec

  Query Match             86.5%;  Score 964.5;  DB 2;  Length 3707;
  Best Local Similarity   85.7%;  Pred. No. 1.7e-70;
  Matches  180;  Conservative   13;  Mismatches   16;  Indels    1;  Gaps    1;

Qy          1 GIAESDWHLEGSGGNDAPGQYGAYFHDDGFLAFPGHVFSRSLPEVPETIELEVRTSTASG 60
              |: ||||| ||||||||||||||||:|:|||  ||: ||||||||||||| ||||||| |
Db       3499 GVVESDWHPEGSGGNDAPGQYGAYFYDNGFLGLPGNSFSRSLPEVPETIEFEVRTSTADG 3558

Qy         61 LLLWQGVEVGEAGQGKDFISLGLQDGHLVFRYQLGSGEARLVSEDPINDGEWHRVTALRE 120
              ||||||| | || : ||||||||||||||| |||||||||||| ||||||||||:|||||
Db       3559 LLLWQGV-VREASRSKDFISLGLQDGHLVFSYQLGSGEARLVSGDPINDGEWHRITALRE 3617

Qy        121 GRRGSIQVDGEELVSGRSPGPNVAVNAKGSVYIGGAPDVATLTGGRFSSGITGCVKNLVL 180
              |:|||||||||:||:||||||||||| |  :|||||||||||| |:||||||||:|||||
Db       3618 GQRGSIQVDGEDLVTGRSPGPNVAVNTKDIIYIGGAPDVATLTRGKFSSGITGCIKNLVL 3677

Qy        181 HSARPGAPPPQPLDLQHRAQAGANTRPCPS 210
              |:||||||||||||||||||||||||||||
Db       3678 HTARPGAPPPQPLDLQHRAQAGANTRPCPS 3707

RESULT 3
T19821
hypothetical protein ZC101.2e - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 23-Sep-2002
C;Accession: T19821; T19819; T19820; T27490; T27488; T27489; T27487; A47648; B47648; C4
R;Baynes, C.
submitted to the EMBL Data Library, March 1997
A;Reference number: Z19182
A;Accession: T19821
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-3375 <WIL>
```

```
A;Cross-references: EMBL:Z93375; PIDN:CAB07569.1; GSPDB:GN00020; CESP:ZC101.2e
A;Experimental source: clone C38C6
A;Accession: T19819
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-2441,'R','KRKH',3369,'GN',3372-3373,'G',3375,'RLRHRRRNAQNGPLSRKTRTTTKLFGSW
A;Cross-references: EMBL:Z93375; PIDN:CAB07567.1; GSPDB:GN00020; CESP:ZC101.2a
A;Experimental source: clone C38C6
A;Accession: T19820
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-1694,'H',1883-2441,'R','KRKH',3369,'GN',3372-3373,'G',3375,'RLRHRRRNAQNGPL
A;Cross-references: EMBL:Z93375; PIDN:CAB07568.1; GSPDB:GN00020; CESP:ZC101.2c
A;Experimental source: clone C38C6
R;Percy, C.
submitted to the EMBL Data Library, March 1997
A;Reference number: Z20375
A;Accession: T27490
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-3375 <WI2>
A;Cross-references: EMBL:Z93395; PIDN:CAB07708.1; GSPDB:GN00020; CESP:ZC101.2e
A;Experimental source: clone ZC101
A;Accession: T27488
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-2441,'R','KRKH',3369,'GN',3372-3373,'G',3375,'RLRHRRRNAQNGPLSRKTRTTTKLFGSW
A;Cross-references: EMBL:Z93395; PIDN:CAB07706.1; GSPDB:GN00020; CESP:ZC101.2a
A;Experimental source: clone ZC101
A;Accession: T27489
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-1694,'H',1883-2441,'R','KRKH',3369,'GN',3372-3373,'G',3375,'RLRHRRRNAQNGPL
A;Cross-references: EMBL:Z93395; PIDN:CAB07707.1; GSPDB:GN00020; CESP:ZC101.2c
A;Experimental source: clone ZC101
A;Accession: T27487
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-1128,1290,'DFARNSPS',1299,'NSS',1303-1304,'R','RHR',1544-1545,'RIRVRS',155
A;Cross-references: EMBL:Z93395; PIDN:CAB07704.1; GSPDB:GN00020; CESP:ZC101.2b
A;Experimental source: clone ZC101
R;Rogalski, T.M.; Williams, B.D.; Mullen, G.P.; Moerman, D.G.
Genes Dev. 7, 1471-1484, 1993
A;Title: Products of the unc-52 gene in Caenorhabditis elegans are homologous to the cor
A;Reference number: A47648; MUID:93339574; PMID:8393416
A;Accession: A47648
A;Status: preliminary; nucleic acid sequence not shown
A;Molecule type: DNA
A;Residues: 1-546,'P',548-2441,'R','KRKH',3369,'GN',3372-3373,'G',3375,'RLRHRRRNAQNGPLSR
A;Cross-references: GB:L13458
A;Accession: B47648
A;Status: preliminary; nucleic acid sequence not shown
A;Molecule type: DNA
A;Residues: 1-546,'P',548-2198,'D',2290,'NAR',2294,'L',2296,'WHATE',2302-2303,'V',2305,'
1,'ANIV',2516-2517,'LQQG',2522,'IDG',2526,'S',2528,'SRGFHV',2535,'F' <RO2>
A;Cross-references: GB:L13458
A;Accession: C47648
A;Status: preliminary; nucleic acid sequence not shown
A;Molecule type: DNA
A;Residues: 1-546,'P',548-1128,1290,'DFARNSPS',1299,'NSS',1303-1304,'R','RHR',1544-1545,
A;Cross-references: GB:L13458
C;Genetics:
A;Gene: CESP:ZC101.2e; CESP:ZC101.2a; CESP:ZC101.2c; CESP:ZC101.2b
A;Map position: 2
A;Introns: 32/1; 134/1; 225/1; 335/2; 450/3; 739/3; 830/3; 860/2; 1064/2; 1129/1; 1158/3
2; 2613/1; 2684/1; 2757/1; 2813/3; 2863/1; 2900/3; 3084/1; 3176/1; 3250/2
C;Superfamily: LR11 protein; laminin-type EGF-like homology; LDL receptor ligand-binding
C;Keywords: extracellular matrix
F;149-183/Domain: LDL receptor ligand-binding repeat homology <LDL1>
F;190-224/Domain: LDL receptor ligand-binding repeat homology <LDL2>
F;233-268/Domain: LDL receptor ligand-binding repeat homology <LDL3>
F;955-1002/Domain: laminin-type EGF-like homology <LEG1>
F;1011-1058/Domain: laminin-type EGF-like homology <LEG2>
```

```
  Query Match             30.4%;  Score 338.5;  DB 2;  Length 3375;
  Best Local Similarity   35.6%;  Pred. No. 3.6e-19;
  Matches   68;  Conservative   38;  Mismatches   76;  Indels    9;  Gaps    2;

Qy          18 PGQYGAYFHDDGFLAFPGHVFSRSLPEVPETIELEVRTSTASGLLLWQGVEVGEAGQGKD 77
               | :: | |  | |:      |     |  | :  : :|   :|:||||| :     | :|
Db        3178 PIEHAARFDGDAFIELSSDEFPHLTSEKDEIVAFKFKTEQQNGVLLWQG-QRPTVQQMED 3236

Qy          78 FISLGLQDGHLVFRYQLGSGEARLVSEDPINDGEWHRVTALREGRRGSIQVDGEELVSGR 137
               :||:|: :||| | |:|| | | |:||: ::||: | |   |:|| | :::|    | ||
Db        3237 YISVGIVNGHLHFSYELGGGAAHLISEERVDDGKEHSVRFERKGREGQMRIDNYREVDGR 3296

Qy         138 SPGPNVAVNAKGSVYIGGAPDVATLTGGRFSSGITGCVKNLVLHSARPGAPPPQPLDLQH 197
               | |    :|  |::::|| ||::  ||| ||:   ||: :: |:  :        |||
Db        3297 STGILAMLNVDGNIFVGGVPDISKATGGLFSNNFVGCIADVELNGVK--------LDLMA 3348

Qy         198 RAQAGANTRPC 208
                |  | | :||
Db        3349 TAIDGKNVKPC 3359


RESULT 4
AGRT
agrin - rat
C;Species: Rattus norvegicus (Norway rat)
C;Date: 31-Mar-1993 #sequence_revision 31-Mar-1993 #text_change 17-Nov-2000
C;Accession: JH0399; A38856
R;Rupp, F.; Payan, D.G.; Magill-Solc, C.; Cowan, D.M.; Scheller, R.H.
Neuron 6, 811-823, 1991
A;Title: Structure and expression of a rat agrin.
A;Reference number: JH0399; MUID:91222570; PMID:1851019
A;Accession: JH0399
A;Molecule type: mRNA
A;Residues: 1-1779;1799-1959 <RUP>
A;Cross-references: GB:M64780; NID:g202798; PIDN:AAA40703.1; PID:g202800
A;Experimental source: embryonic spinal cord
A;Note: it is uncertain whether Met-1, Met-18, or Met-24 is the initiator
R;Rupp, F.; Oezcelik, T.; Linial, M.; Peterson, K.; Francke, U.; Scheller, R.
J. Neurosci. 12, 3535-3544, 1992
A;Title: Structure and chromosomal localization of the mammalian agrin gene.
A;Reference number: A38856; MUID:92407628; PMID:1326608
A;Accession: A38856
A;Molecule type: mRNA
A;Residues: 1780-1798 <RU2>
A;Cross-references: GB:S44194
C;Comment: This protein mediates the motor neuron-induced aggregation of acetylcholine r
C;Comment: 90% of rat embryonic transcripts encode the variant labeled below as form 3.
ylcholine receptor clustering activity.
C;Superfamily: agrin; EGF homology; Kazal proteinase inhibitor homology; laminin G repea
C;Keywords: alternative splicing; duplication; glycoprotein; neuromuscular junction
F;1-1959/Product: agrin, form 1 #status predicted <AG1>
F;1-1787,1799-1959/Product: agrin, form 4 #status predicted <AG4>
F;1-1779,1799-1959/Product: agrin, form 3 #status predicted <AG3>
F;1-1779,1788-1959/Product: agrin, form 5 #status predicted <AG5>
F;1-1143,1153-1959/Product: agrin, form 2 #status predicted <AG2>
F;22-50/Region: hydrophobic
F;88-137/Domain: Kazal proteinase inhibitor homology <KPI1>
F;163-212/Domain: Kazal proteinase inhibitor homology <KPI2>
F;236-284/Domain: Kazal proteinase inhibitor homology <KPI3>
F;307-356/Domain: Kazal proteinase inhibitor homology <KPI4>
F;381-429/Domain: Kazal proteinase inhibitor homology <KPI5>
F;446-494/Domain: Kazal proteinase inhibitor homology <KPI6>
F;511-559/Domain: Kazal proteinase inhibitor homology <KPI7>
F;540-542/Region: motor neuron attachment (L-R-E) motif
F;596-645/Domain: Kazal proteinase inhibitor homology <KPI8>
F;688-739/Domain: laminin-type EGF-like homology <LE1>
F;742-786/Domain: laminin-type EGF-like homology <LE2>
F;814-864/Domain: Kazal proteinase inhibitor homology <KPI9>
F;869-992/Region: serine/threonine-rich
F;1084-1086/Region: motor neuron attachment (L-R-E) motif
F;1147-1215/Region: serine/threonine-rich
F;1224-1257/Domain: EGF homology <EG1>
```

```
F;1287-1442/Domain: laminin G repeat homology <LG1>
F;1444-1476/Domain: EGF homology <EG2>
F;1483-1515/Domain: EGF homology <EG3>
F;1555-1706/Domain: laminin G repeat homology <LG2>
F;1713-1747/Domain: EGF homology <EG4>
F;1807-1959/Domain: laminin G repeat homology <LG3>
F;97-116,105-137,171-191,180-212,244-263,252-284,316-335,324-356,389-408,397-429,454-473
-1476,1483-1494,1488-1504,1506-1515/Disulfide bonds: #status predicted
F;145,672,827,957/Binding site: carbohydrate (Asn) (covalent) #status predicted

  Query Match             21.7%;  Score 242.5;  DB 1;  Length 1959;
  Best Local Similarity   32.6%;  Pred. No. 1.4e-11;
  Matches   61;  Conservative   26;  Mismatches   87;  Indels   13;  Gaps    3;

Qy          25 FHDDGFLAFPGHVFSRSLPEVPETIELEVRTSTASGLLLWQGVEVGEAGQGKDFISLGLQ 84
               |    |||||      :|      : || |     ||||: |       :||||::| |
Db        1268 FKGHSFLAFPTLRAYHTL-----RLALEFRALETEGLLLYNG-----NARGKDFLALALL 1317

Qy          85 DGHLVFRYQLGSGEARLVSEDPINDGEWHRVTALREGRRGSIQVDGEELVSGRSPGPNVA 144
               || : ||:  ||| | | |  |:  | |||:   |  |:|:: ||||  | | ||
Db        1318 DGRVQFRFDTGSGPAVLTSLVPVEPGRWHRLELSRHWRQGTLSVDGETPVVGESPSGTDG 1377

Qy         145 VNAKGSVYIGGAPD---VATLTGGRFSSGITGCVKNLVLHSARPGAPPPQPLDLQHRAQA 201
               :|   ::|:|| |:      |       |: ||:: | ::: :      |   :|
Db        1378 LNLDTNLYVGGIPEEQVAMVLDRTSVGVGLKGCIRMLDINNQQLELSDWQRAAVQSSGVG 1437

Qy         202 GANTRPC 208
                    ||
Db        1438 ECGDHPC 1444

RESULT 5
T43060
agrin - electric ray (Discopyge ommata) (fragment)
C;Species: Discopyge ommata
C;Date: 11-Jan-2000 #sequence_revision 11-Jan-2000 #text_change 17-Nov-2000
C;Accession: T43060
R;Smith, M.A.; Magill-Solc, C.; Rupp, F.; Yao, Y.M.M.; Schilling, J.W.; Snow, P.; McMaha
submitted to the EMBL Data Library, September 1992
A;Reference number: Z22308
A;Accession: T43060
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-1328 <SMI>
A;Cross-references: EMBL:L01423; NID:g213102; PID:g213103; PIDN:AAA49224.1
C;Superfamily: agrin; EGF homology; Kazal proteinase inhibitor homology; laminin G repea
C;Keywords: glycoprotein; neuromuscular junction

  Query Match             21.1%;  Score 235;  DB 2;  Length 1328;
  Best Local Similarity   36.9%;  Pred. No. 3.7e-11;
  Matches   62;  Conservative   32;  Mismatches   56;  Indels   18;  Gaps    8;

Qy          25 FHDDGFLAFPG-HVFSRSLPEVPETIELEV--RTSTASGLLLWQGVEVGEAGQGKDFISL 81
               |:   :|   | | |   : :: : : :||       :|:: : | :    |:| ||:||
Db         919 FNGLSYLEMNGIHTF---VSDLLQKLSMEVIFLAKDPNGMIFYNGQKTD--GRG-DFVSL 972

Qy          82 GLQDGHLVFRYQLGSGEARLVSEDPINDGEWHRVTALREGRRGSIQVDGEELVSGRSP-- 139
                |:||:| |:| || | | | |: ||    |: ||  | ||:| :::: :||||| ||
Db         973 NLRDGYLEFKYDLGKGAAVLRSKAPIPLNVWNVVTVERNGRKGLMKINKDELVSGESPKS 1032

Qy         140 --GPNVAVNAKGSVYIGGAPD---VATLTGGRFSSGITGCVKNLVLHS 182
                  |: |:| | : |:|||||    |   |    || || :: | | |
Db        1033 RKAPHTALNLKEAFYVGGAPDFNKFARAAG--IISGFTGAIQKLSLKS 1078

RESULT 6
MMHUMH
laminin alpha-2 chain - human (fragment)
N;Alternate names: laminin M chain; merosin heavy chain
C;Species: Homo sapiens (man)
C;Date: 30-Sep-1991 #sequence_revision 21-Aug-1998 #text_change 10-Dec-1999
C;Accession: PX0082; A35899; A38970; S14461
R;Hori, H.; Kanamori, T.; Mizuta, T.; Yamaguchi, N.; Liu, Y.; Nagai, Y.
```

```
J. Biochem. 116, 1212-1219, 1994
A;Title: Human laminin M chain: Epitope analysis of its monoclonal antibodies by immunos
A;Reference number: PX0082; MUID:95221315; PMID:7535762
A;Accession: PX0082
A;Molecule type: mRNA
A;Residues: 1-1751 <HOR>
A;Experimental source: placenta
R;Ehrig, K.; Leivo, I.; Argraves, W.S.; Ruoslahti, E.; Engvall, E.
Proc. Natl. Acad. Sci. U.S.A. 87, 3264-3268, 1990
A;Title: Merosin, a tissue-specific basement membrane protein, is a laminin-like protei
A;Reference number: A35899; MUID:90238994; PMID:2185464
A;Accession: A35899
A;Status: nucleic acid sequence not shown
A;Molecule type: mRNA
A;Residues: 'V',623-1751 <EHR1>
A;Cross-references: EMBL:M59832
A;Accession: A38970
A;Molecule type: protein
A;Residues: 1368-1384;1389-1406;1593-1607 <EHR2>
A;Note: the sequence from Fig. 1 is inconsistent with that from Fig. 2 in lacking 1599-
R;Ehrig, K.; Leivo, I.; Argraves, S.W.; Ruoslahti, E.; Engvall, E.
submitted to the EMBL Data Library, December 1990
A;Description: The tissue-specific basement membrane protein merosin is a laminin-like
A;Reference number: S14461
A;Accession: S14461
A;Molecule type: mRNA
A;Residues: 'V',623-1264,'R',1266-1751 <LEI>
A;Cross-references: EMBL:M59832; NID:g187520; PIDN:AAA63215.1; PID:g187521
C;Comment: This protein is a prominent component of the basement membrane that mediates
C;Genetics:
A;Gene: GDB:LAMA2; LAMM
A;Cross-references: GDB:132362; OMIM:156225
A;Map position: 6q22-6q23
C;Complex: Laminins are trimers of an alpha-type, a beta-type, and a gamma-type laminin
C;Function:
A;Description: interact with cells and with other basement membrane proteins to promote
C;Superfamily: laminin alpha-1 chain; laminin G repeat homology; laminin-type EGF-like
C;Keywords: basement membrane; calcium binding; coiled coil; extracellular matrix; glyc
F;21-58/Domain: laminin-type EGF-like homology #status atypical <LE01>
F;61-107/Domain: laminin-type EGF-like homology <LE02>
F;110-165/Domain: laminin-type EGF-like homology <LE03>
F;168-212/Domain: laminin-type EGF-like homology <LE04>
F;527-567,1071-1300/Region: 3DM and 2D9 binding
F;811-972/Domain: laminin G repeat homology <LG1>
F;1005-1165/Domain: laminin G repeat homology <LG2>
F;1191-1354/Domain: laminin G repeat homology <LG3>
F;1430-1578/Domain: laminin G repeat homology <LG4>
F;1605-1751/Domain: laminin G repeat homology <LG5>
F;120,238,255,341,451,542,557,561,658,669,686,767,881,1001,1076,1119,1192,1199,1289,150

  Query Match             20.0%;  Score 223.5;  DB 1;  Length 1751;
  Best Local Similarity   36.4%;  Pred. No. 4.5e-10;
  Matches   59;  Conservative   23;  Mismatches   67;  Indels   13;  Gaps    6;

Qy          48 TIELEVRTSTASGLLLWQGVEVGEAGQGKDFISLGLQDGHLVFRYQLGSGEARLVSEDPI 107
               ||||||||   |||| :    :  |    || :: |::|   | | ||||:   :    |
Db        1429 TIELEVRTEAESGLLFYM-ARINHA----DFATVQLRNGLPYFSYDLGSGDTHTMIPTKI 1483

Qy         108 NDGEWHRVTALREGRRGSIQVDGEELVSGRSPGPNVA--VNAKGSVYIGGAP-DVATLTG 164
               |||:||::  :|  : | : |||    | |:  |  |  ::  | :|:|| | :  |
Db        1484 NDGQWHKIKIMRSKQEGILYVDG---ASNRTISPKKADILDVVGMLYVGGLPINYTTRRI 1540

Qy         165 GRFSSGITGCVKNLVLHSARPGAPPPQPLDLQHRAQAGANTR 206
               |  :  | |||:|  || |   |   ||    |     || :
Db        1541 GPVTYSIDGCVRN--LHMAEAPADLEQPTSSFHVGTCFANAQ 1580

RESULT 7
S53868
laminin alpha-2 chain precursor - mouse
N;Alternate names: laminin M chain; merosin heavy chain
C;Species: Mus musculus (house mouse)
C;Date: 21-Aug-1998 #sequence_revision 21-Aug-1998 #text_change 10-Dec-1999
```

```
C;Accession: I49077; S50829; I48655; S31576; S53868
R;Bernier, S.M.; Utani, A.; Sugiyama, S.; Doi, T.; Polistina, C.; Yamada, Y.
Matrix Biol. 14, 447-455, 1995
A;Title: Cloning and expression of laminin alpha 2 chain (M-chain) in the mouse.
A;Reference number: I49077; MUID:95316259; PMID:7795883
A;Accession: I49077
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-3106 <RES>
A;Cross-references: EMBL:U12147; NID:g699109; PIDN:AAC52165.1; PID:g699110
R;Xu, H.; Wu, X.R.; Wewer, U.M.; Engvall, E.
Nature Genet. 8, 297-302, 1994
A;Title: Murine muscular dystrophy caused by a mutation in the laminin alpha-2 (Lama2) g
A;Reference number: S50829; MUID:95179178; PMID:7874173
A;Accession: S50829
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 64-281 <XUH>
A;Cross-references: GB:S75315; NID:g833929; PIDN:AAB33573.1; PID:g833930
R;Chang, A.C.; Wadsworth, S.; Coligan, J.E.
J. Immunol. 151, 1789-1801, 1993
A;Title: Expression of merosin in the thymus and its interaction with thymocytes.
A;Reference number: I48655; MUID:93346725; PMID:8345183
A;Accession: I48655
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 2162-2204,'D',2206-2213,'EY',2216-2279 <RE2>
A;Cross-references: EMBL:X69869; NID:g53055; PIDN:CAA49502.1; PID:g53056
C;Complex: Laminins are trimers of an alpha-type, a beta-type, and a gamma-type laminin
C;Function:
A;Description: interact with cells and with other basement membrane proteins to promote
C;Superfamily: laminin alpha-1 chain; laminin G repeat homology; laminin-type EGF-like h
C;Keywords: basement membrane; calcium binding; coiled coil; extracellular matrix; glyc
F;1-22/Domain: signal sequence #status predicted <SIG>
F;23-3106/Product: laminin alpha-2 chain #status predicted <MAT>
F;283-337/Domain: laminin-type EGF-like homology <LE01>
F;340-407/Domain: laminin-type EGF-like homology <LE02>
F;410-462/Domain: laminin-type EGF-like homology <LE03>
F;465-511/Domain: laminin-type EGF-like homology <LE04>
F;514-523/Domain: laminin-type EGF-like homology #status atypical <LE05>
F;720-750/Domain: laminin-type EGF-like homology <LE06>
F;753-800/Domain: laminin-type EGF-like homology <LE07>
F;803-858/Domain: laminin-type EGF-like homology <LE08>
F;861-911/Domain: laminin-type EGF-like homology <LE09>
F;914-960/Domain: laminin-type EGF-like homology <LE10>
F;963-1007/Domain: laminin-type EGF-like homology <LE11>
F;1010-1053/Domain: laminin-type EGF-like homology <LE12>
F;1056-1099/Domain: laminin-type EGF-like homology <LE13>
F;1102-1121/Domain: laminin-type EGF-like homology #status atypical <LE14>
F;1123-1159/Domain: laminin-type EGF-like homology #status atypical <LE15>
F;1162-1171/Domain: laminin-type EGF-like homology #status atypical <LE16>
F;1376-1413/Domain: laminin-type EGF-like homology #status atypical <LE17>
F;1416-1462/Domain: laminin-type EGF-like homology <LE18>
F;1465-1520/Domain: laminin-type EGF-like homology <LE19>
F;1523-1567/Domain: laminin-type EGF-like homology <LE20>
F;2166-2327/Domain: laminin G repeat homology <LG1>
F;2360-2520/Domain: laminin G repeat homology <LG2>
F;2546-2709/Domain: laminin G repeat homology <LG3>
F;2785-2933/Domain: laminin G repeat homology <LG4>
F;2960-3106/Domain: laminin G repeat homology <LG5>

  Query Match             18.6%;  Score 207.5;  DB 1;  Length 3106;
  Best Local Similarity   35.8%;  Pred. No. 1.8e-08;
  Matches   54;  Conservative   25;  Mismatches   53;  Indels   19;  Gaps    6;

Qy         48 TIELEVRTSTASGLLLWQGVEVGEAGQGKDFISLGLQDGHLVFRYQLGSGEARLVSEDPI 107
              ||||||||   |||| : |  :  |    || :: |::|   | | ||||  | :    |
Db       2784 TIELEVRTEAESGLLFYMG-RINHA----DFGTVQLRNGFPFFSYDLGSGSTRTMIPTKI 2838

Qy        108 NDGEWHRVTALREGRRGSIQVDGEELVSGRSPGPNVA--VNAKGSVYIGGAP-DVATLTG 164
              |||:||::  :|  : | : ||     | ::  |  |  ::  | :|:|| | :  |
Db       2839 NDGQWHKIKIVRVKQEGILYVDD---ASSQTISPKKADILDVGGILYVGGLPINYTTRRI 2895

Qy        165 GRFSSGITGCVKNLVLHSARPGAPPPQPLDL 195
              |  :  : |||:|| :  |        |:||
Db       2896 GPVTYSLDGCVRNLHMEQA--------PVDL 2918


RESULT 8
AGCH
agrin precursor - chicken
C;Species: Gallus gallus (chicken)
C;Date: 31-Mar-1993 #sequence_revision 31-Mar-1993 #text_change 17-Nov-2000
C;Accession: JH0591; A38857; B38857; I50692
R;Tsim, K.W.K.; Ruegg, M.A.; Escher, G.; Kroeger, S.; McMahan, U.J.
Neuron 8, 677-689, 1992
A;Title: cDNA that encodes active agrin.
A;Reference number: JH0591; MUID:92232297; PMID:1314620
A;Accession: JH0591
A;Molecule type: mRNA
A;Residues: 1-1955 <TSI>
A;Cross-references: GB:M94271; NID:g211120; PIDN:AAA48585.1; PID:g211121
A;Experimental source: brain
R;Ruegg, M.A.; Tsim, K.W.K.; Horton, S.E.; Kroeger, S.; Escher, G.; Gensch, E.M.; McMah
Neuron 8, 691-699, 1992
A;Title: The agrin gene codes for a family of basal lamina proteins that differ in funct
A;Reference number: A38857; MUID:92232298; PMID:1314621
A;Contents: alternative splicing
A;Accession: A38857
A;Molecule type: mRNA
A;Residues: 1132-1783;1795-1955 <RU2>
A;Cross-references: GB:M97371
A;Accession: B38857
A;Molecule type: mRNA
A;Residues: 1221-1647;1652-1783;1794-1955 <RU3>
A;Cross-references: GB:M97372
A;Note: translation of the nucleotide sequence is not complete
R;Thomas, W.S.; O'Dowd, D.K.; Smith, M.A.
Dev. Biol. 158, 523-535, 1993
A;Title: Developmental expression and alternative splicing of chick agrin RNA.
A;Reference number: I50692; MUID:93345745; PMID:8393816
A;Accession: I50692
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 'SHLSNEIPA',1784-1795 <THO>
A;Cross-references: EMBL:U07271; NID:g459665; PIDN:AAA16788.1; PID:g459666
C;Comment: This protein mediates the motor neuron-induced aggregation of acetylcholine
C;Comment: Alternative splicing produces two inactive proteins: agrin-related protein 1
C;Superfamily: agrin; EGF homology; Kazal proteinase inhibitor homology; laminin G repe
C;Keywords: alternative splicing; duplication; glycoprotein; neuromuscular junction
F;1-38/Domain: signal sequence #status predicted <SIG>
F;39-1955/Product: agrin #status predicted <MAT>
F;39-1783,1795-1955/Product: agrin-related protein 1 #status predicted <AG1>
F;39-1647,1652-1783,1794-1955/Product: agrin-related protein 2 #status predicted <AG2>
F;77-126/Domain: Kazal proteinase inhibitor homology <KPI1>
F;152-201/Domain: Kazal proteinase inhibitor homology <KPI2>
F;225-273/Domain: Kazal proteinase inhibitor homology <KPI3>
F;295-344/Domain: Kazal proteinase inhibitor homology <KPI4>
F;370-418/Domain: Kazal proteinase inhibitor homology <KPI5>
F;435-483/Domain: Kazal proteinase inhibitor homology <KPI6>
F;500-548/Domain: Kazal proteinase inhibitor homology <KPI7>
F;584-633/Domain: Kazal proteinase inhibitor homology <KPI8>
F;675-726/Domain: laminin-type EGF-like homology <LE1>
F;729-773/Domain: laminin-type EGF-like homology <LE2>
F;801-851/Domain: Kazal proteinase inhibitor homology <KPI9>
F;856-995/Region: serine/threonine-rich
F;1150-1219/Region: serine/threonine-rich
F;1233-1264/Domain: EGF homology <EG1>
F;1294-1448/Domain: laminin G repeat homology <LG1>
F;1429-1431/Region: motor neuron attachment (L-R-E) motif
F;1450-1482/Domain: EGF homology <EG2>
F;1489-1521/Domain: EGF homology <EG3>
F;1560-1711/Domain: laminin G repeat homology <LG2>
F;1718-1751/Domain: EGF homology <EG4>
F;1803-1955/Domain: laminin G repeat homology <LG3>
F;86-105,94-126,160-180,169-201,233-252,241-273,304-323,312-344,378-397,386-418,443-462,
```

```
1482,1489-1500,1494-1510,1512-1521/Disulfide bonds: #status predicted
F;390,659,764,814/Binding site: carbohydrate (Asn) (covalent) #status predicted

  Query Match             18.0%;  Score 200.5;  DB 1;  Length 1955;
  Best Local Similarity   32.9%;  Pred. No. 3.9e-08;
  Matches   56;  Conservative   30;  Mismatches   65;  Indels   19;  Gaps    8;

Qy         48 TIELEVRTSTASGLLLWQGVEVGEAGQGKDFISLGLQDGHLVFRYQLGSGEARLVSEDPI 107
              ::|:     : ||:: : | :    |:| ||:|| | ||:| :|| || | | | |::|:
Db       1559 SMEVVFLAKSPSGMIFYNGQKTD--GKG-DFVSLALHDGYLEYRYDLGKGAAVLRSKEPV 1615

Qy        108 NDGEWHRVTALREGRRGSIQVDGEELVSGRSPG----PNVAVNAKGSVYIGGAPDVATLT 163
                  |  |   | ||:| ::::  | | | ||     |:  :| |   |:||||| : |
Db       1616 PLNTWISVLLERSGRKGVMRINNGERVMGESPKSRKVPHAFLNLKEPFYVGGAPDFSKLA 1675

Qy        164 -GGRFSSGITGCVKNLVLHSARPGAPPPQPLDLQH-RAQAGANT---RPC 208
                   |:   | |: : :     | |    |  || |:    :|    ||
Db       1676 RAAAISTSFYGAVQRISI----KGVP---LLKEQHIRSAVEISTFRAHPC 1718


RESULT 9
S14458
laminin alpha-1 chain precursor - human
C;Species: Homo sapiens (man)
C;Date: 30-Sep-1991 #sequence_revision 30-Sep-1991 #text_change 19-Jan-2001
C;Accession: S14458; S14663; A34961
R;Haaparanta, T.; Uitto, J.; Ruoslahti, E.; Engvall, E.
Matrix 11, 151-160, 1991
A;Title: Molecular cloning of the cDNA encoding human laminin A chain.
A;Reference number: S14458; MUID:91333420; PMID:1714537
A;Accession: S14458
A;Status: not compared with conceptual translation
A;Molecule type: mRNA
A;Residues: 1-3075 <HAA>
R;Nissinen, M.; Vuolteenaho, R.; Boot-Handford, R.; Kallunki, P.; Tryggvason, K.
Biochem. J. 276, 369-379, 1991
A;Title: Primary structure of the human laminin A chain. Limited expression in human tis
A;Reference number: S14663; MUID:91264789; PMID:2049067
A;Accession: S14663
A;Molecule type: mRNA
A;Residues: 1-227,'FE',230-251,'MLP',255-418,'E',420-518,'L',520-1022,'V',1024-1074,'V',
A;Cross-references: EMBL:X58531; NID:g34225; PIDN:CAA41418.1; PID:g34226
R;Olsen, D.; Nagayoshi, T.; Fazio, M.; Peltonen, J.; Jaakkola, S.; Sanborn, D.; Sasaki,
Lab. Invest. 60, 772-782, 1989
A;Title: Human laminin: cloning and sequence analysis of cDNAs encoding A, B1 and B2 cha
A;Reference number: A34961; MUID:89280632; PMID:2733383
A;Accession: A34961
A;Status: not compared with conceptual translation
A;Molecule type: mRNA
A;Residues: 'W',2397-2745,'L',2747-3053,'L',3055-3072,'PSP' <OLS>
A;Note: the authors translated the codon AGA for residue 2692 as Pro
C;Genetics:
A;Gene: GDB:LAMA1; LAMA
A;Cross-references: GDB:120135; OMIM:150320
A;Map position: 18p11.32-18p11.22
C;Superfamily: laminin alpha-1 chain; laminin G repeat homology; laminin-type EGF-like h
C;Keywords: basement membrane; calcium binding; cell binding; coiled coil; disulfide bon
F;1-17/Domain: signal sequence #status predicted <SIG>
F;18-3075/Product: laminin alpha-1 chain #status predicted <MAT>
F;18-269/Domain: VI <DOM6>
F;270-516/Domain: V <DOM5>
F;270-324/Domain: laminin-type EGF-like homology <LE1>
F;327-394/Domain: laminin-type EGF-like homology <LE2>
F;397-451/Domain: laminin-type EGF-like homology <LE3>
F;454-500/Domain: laminin-type EGF-like homology <LE4>
F;503-512/Domain: laminin-type EGF-like homology #status atypical <LE5>
F;517-708/Domain: IVb <DO4B>
F;709-1159/Domain: IIIb <DO3B>
F;709-739/Domain: laminin-type EGF-like homology #status atypical <LE6>
F;742-788/Domain: laminin-type EGF-like homology <LE7>
F;791-846/Domain: laminin-type EGF-like homology <LE8>
F;849-899/Domain: laminin-type EGF-like homology <LE9>
F;902-948/Domain: laminin-type EGF-like homology <LE10>
```

```
F;951-995/Domain: laminin-type EGF-like homology <LE11>
F;998-1041/Domain: laminin-type EGF-like homology <LE12>
F;1044-1087/Domain: laminin-type EGF-like homology <LE13>
F;1090-1109/Domain: laminin-type EGF-like homology #status atypical <LE14>
F;1111-1147/Domain: laminin-type EGF-like homology #status atypical <LE15>
F;1150-1159/Domain: laminin-type EGF-like homology #status atypical <LE16>
F;1160-1361/Domain: IVa <DO4A>
F;1362-1553/Domain: IIIa <DO3A>
F;1362-1400/Domain: laminin-type EGF-like homology #status atypical <LE17>
F;1403-1449/Domain: laminin-type EGF-like homology <LE18>
F;1452-1506/Domain: laminin-type EGF-like homology <LE19>
F;1509-1553/Domain: laminin-type EGF-like homology <LE20>
F;1554-2125/Domain: I/II, heptad repeats <DOM2>
F;2116-2120/Region: cell adhesion #status predicted
F;2126-3075/Domain: G <DOMG>
F;2142-2300/Domain: laminin G repeat homology <LG1>
F;2329-2484/Domain: laminin G repeat homology <LG2>
F;2510-2676/Domain: laminin G repeat homology <LG3>
F;2534-2536/Region: cell attachment (R-G-D) motif
F;2739-2888/Domain: laminin G repeat homology <LG4>
F;2916-3073/Domain: laminin G repeat homology <LG5>
F;38,164,555,665,763,801,838,926,952,1045,1407,1579,1596,1678,1689,1698,1717,1804,1894,
rate (Asn) (covalent) #status predicted
F;297-305/Disulfide bonds: #status predicted

  Query Match             16.9%;  Score 188.5;  DB 2;  Length 3075;
  Best Local Similarity   30.5%;  Pred. No. 6.6e-07;
  Matches   54;  Conservative   30;  Mismatches   70;  Indels   23;  Gaps    7;

Qy         11 GSGGNDAPGQYGAYFHDDGFLAFPGH-VFSRSLPEVPETIELEVRTSTASGLLLWQGVEV 69
              ||  |: |      || ||     |: |  :|||     | :   | : :||||:    :
Db       2299 GSSQNEDPS-----FHFDG----SGYSVVEKSLPATVTQIIMLFNTFSPNGLLLY----L 2345

Qy         70 GEAGQGKDFISLGLQDGHLVFRYQLGSGEARLVSEDPINDGEWHRVTALREGRRGSIQV- 128
              |  |  |||:|: |  | :     ||||   |:::   |:| |:::   |  ::| : |
Db       2346 GSYGT-KDFLSIELFRGRVKVMTDLGSGPITLLTDRRYNNGTWYKIAFQRNRKQGVLAVI 2404

Qy        129 -----DGEELVSGRSPGPNVAVNA--KGSVYIGGAPDVATLTGGRFSSGITGCVKNL 178
                     :|   | :|| :  :|   |  :|:|| |    :  |  :    ||:|||
Db       2405 DAYNTSNKETKQGETPGASSDLNRLDKDPIYVGGLPRSRVVRRGVTTKSFVGCIKNL 2461


RESULT 10
C40228
neurexin II-alpha precursor - rat
C;Species: Rattus norvegicus (Norway rat)
C;Date: 21-Jul-1995 #sequence_revision 21-Jul-1995 #text_change 09-Dec-2002
C;Accession: C40228; S27886; S27887
R;Ushkaryov, Y.A.; Petrenko, A.G.; Geppert, M.; Suedhof, T.C.
Science 257, 50-56, 1992
A;Title: Neurexins: synaptic cell surface proteins related to the alpha-latrotoxin rece
A;Reference number: A40228; MUID:92320296; PMID:1621094
A;Accession: C40228
A;Status: preliminary; nucleic acid sequence not shown
A;Molecule type: mRNA
A;Residues: 1-1715 <USH>
A;Cross-references: GB:M96376; NID:g205714; PIDN:AAA41707.1; PID:g205716
A;Note: authors translated the codon GAC for residue 1464 as Thr and ACC for residue 14
R;Ushkaryov, Y.A.; Petrenko, A.G.; Geppert, M.; Sudhof, T.C.
submitted to the EMBL Data Library, July 1992
A;Description: Neurexins: Synaptic cell surface proteins related to the alpha-lotrotoxi
A;Reference number: S27884
A;Accession: S27886
A;Molecule type: mRNA
A;Residues: 1-1666,'CRK',1670,'PREEKLLPG',1683-1685,'GL',1688,'LDLA',1694-1695,'CCVCRCR
A;Cross-references: EMBL:M96376; NID:g205714; PIDN:AAA41706.1; PID:g205715
C;Genetics:
A;Introns: 1666/2
C;Superfamily: neurexin; EGF homology
C;Keywords: alternative splicing; transmembrane protein
F;1-27/Domain: signal sequence #status predicted <SIG>
F;28-1715/Product: neurexin II-alpha #status predicted <MAT>
F;694-726/Domain: EGF homology <EGF>
```

```
F;1103-1135/Domain: EGF homology <EGF1>

  Query Match             16.9%;  Score 188;  DB 2;  Length 1715;
  Best Local Similarity   29.9%;  Pred. No. 3.6e-07;
  Matches   53;  Conservative   30;  Mismatches   56;  Indels   38;  Gaps    7;

Qy          28 DGFLAFPGHVFSRSLPEVPETIELEVRTSTASGLLLWQ-------GVEVGEAGQGKDFIS 80
               : |:| |     |:      :| |: ||:  :||||:        ||    : |  |: :
Db         501 EAFVALPRWSAKRT-----GSISLDFRTTEPNGLLLFSQGRRAGAGVGSHSSSQRADYFA 555

Qy          81 LGLQDGHLVFRYQLGSGEARL-VSEDPINDGEWHRVTALREGRRGSIQVDGEE---LVSG 136
               : | ||:|     :|||  :|  |   :|||||  |   |:||:||| |:      | :|
Db         556 MELLDGYLYLLLDMGSGGIKLRASSRKVNDGEWCHVDFQRDGRKGSISVNSRSTPFLATG 615

Qy         137 RSPGPNVAVNAKGSVYIGGAPDVATLTGGR-------------FSSGITGCVKNLVL 180
                |      :: :  :|:|| |:     |||               :|  |||::| :
Db         616 ESE----VLDLESELYLGGLPE-----GGRVDLPLPPEVWTAALRAGYVGCVRDLFI 663

RESULT 11
T23433
hypothetical protein K08C7.3 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 31-Jan-2000
C;Accession: T23433
R;Berks, M.
submitted to the EMBL Data Library, March 1996
A;Reference number: Z19740
A;Accession: T23433
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-3672 <WIL>
A;Cross-references: EMBL:Z70286; PIDN:CAA94293.1; GSPDB:GN00022; CESP:K08C7.3
A;Experimental source: clone K08C7
C;Genetics:
A;Gene: CESP:K08C7.3
A;Map position: 4
A;Introns: 66/1; 284/3; 563/1; 1187/3; 1248/3; 1300/1; 1460/1; 1623/3; 2361/3; 2988/3; 3
C;Superfamily: laminin alpha-1 chain; laminin G repeat homology; laminin-type EGF-like

  Query Match             16.6%;  Score 185.5;  DB 2;  Length 3672;
  Best Local Similarity   29.0%;  Pred. No. 1.4e-06;
  Matches   54;  Conservative   30;  Mismatches   55;  Indels   47;  Gaps    8;

Qy          28 DGFLAF-PGHVFSRSLPEVPETIELEVRTSTASGLLLWQGVEVGEAGQGKDFISLGLQDG 86
               :|: :: | |   |   :    | |   | :  ||| :        |: |||::| | ||
Db        2904 EGYTSYKPSHWNPRKATK----ISLSFLTFSPHGLLFF-------VGKDKDFMALELSDG 2952

Qy          87 HLVFRYQLGSGEARLVSEDP-INDGEWHRVTALREGRRGSIQVDGE-ELVSGRSPGPNVA 144
                :     ||||  : ::|    |||:|| |: :|| :   | :||| |:: |  || :
Db        2953 GVKLSVDLGSGVGQWITESSNYNDGKWHTVSIVREEKHVKIMIDGETEVLEGDVPGKDSE 3012

Qy         145 VNAKGSVYIGGAPDVATLTGGRFSSGIT---------GCVKNLVLHS------------- 182
               ::    :|||| |           ||::         ||:|:: | |
Db        3013 MSVTEFLYIGGTP-----------SGLSVRTTIVPLRGCIKSVKLGSDNVDLESSHASKG 3061

Qy         183 ARPGAP 188
                | | |
Db        3062 VRSGCP 3067

RESULT 12
T37316
probable laminin alpha chain - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 31-Jan-2000
C;Accession: T37316
R;Joh, K.; Zhu, K.; Hedgecock, E.M.; Inoue, T.; Hori, K.
submitted to the EMBL Data Library, August 1998
A;Description: Laminin alpha chain gene in the nematode C. elegans.
A;Reference number: Z21681
A;Accession: T37316
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-3704 <JOH>
A;Cross-references: EMBL:AB016806; PIDN:BAA32347.1
A;Experimental source: strain N2
C;Genetics:
A;Gene: epi-1
A;Map position: IV
A;Introns: 66/1; 284/3; 563/1; 1187/3; 1248/3; 1300/1; 1460/1; 1623/3; 2361/3; 2988/3;
C;Superfamily: laminin alpha-1 chain; laminin G repeat homology; laminin-type EGF-like

  Query Match             16.6%;  Score 185.5;  DB 2;  Length 3704;
  Best Local Similarity   29.0%;  Pred. No. 1.4e-06;
  Matches   54;  Conservative   30;  Mismatches   55;  Indels   47;  Gaps    8;

Qy          28 DGFLAF-PGHVFSRSLPEVPETIELEVRTSTASGLLLWQGVEVGEAGQGKDFISLGLQDG 86
               :|: :: | |   |   :    | |   | :  ||| :        |: |||::| | ||
Db        2904 EGYTSYKPSHWNPRKATK----ISLSFLTFSPHGLLFF-------VGKDKDFMALELSDG 2952

Qy          87 HLVFRYQLGSGEARLVSEDP-INDGEWHRVTALREGRRGSIQVDGE-ELVSGRSPGPNVA 144
                :     ||||  : ::|    |||:|| |: :|| :   | :||| |:: |  || :
Db        2953 GVKLSVDLGSGVGQWITESSNYNDGKWHTVSIVREEKHVKIMIDGETEVLEGDVPGKDSE 3012

Qy         145 VNAKGSVYIGGAPDVATLTGGRFSSGIT---------GCVKNLVLHS------------- 182
               ::    :|||| |           ||::         ||:|:: | |
Db        3013 MSVTEFLYIGGTP-----------SGLSVRTTIVPLRGCIKSVKLGSDNVDLESSHASKG 3061

Qy         183 ARPGAP 188
                | | |
Db        3062 VRSGCP 3067

RESULT 13
T20721
hypothetical protein F25F2.2 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 19-May-2000
C;Accession: T20721; T21343; T23842
R;Ainscough, R.
submitted to the EMBL Data Library, August 1994
A;Reference number: Z19314
A;Accession: T20721
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-4307 <WIL>
A;Cross-references: EMBL:Z35662; PIDN:CAA84721.1; GSPDB:GN00021; CESP:F25F2.2
A;Experimental source: clone F10G11
R;Ainscough, R.
submitted to the EMBL Data Library, August 1994
A;Reference number: Z19410
A;Accession: T21343
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-4307 <WI2>
A;Cross-references: EMBL:Z35599; PIDN:CAA84661.1; GSPDB:GN00021; CESP:F25F2.2
A;Experimental source: clone F25F2
R;Sulston, J.
submitted to the EMBL Data Library, June 1994
A;Reference number: Z19806
A;Accession: T23842
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-4307 <WI3>
A;Cross-references: EMBL:Z34802; PIDN:CAA84339.1; GSPDB:GN00021; CESP:F25F2.2
A;Experimental source: clone M88
C;Genetics:
A;Gene: CESP:F25F2.2
A;Map position: 3
A;Introns: 79/3; 167/3; 251/3; 328/3; 587/3; 625/3; 737/3; 1006/1; 1190/2; 1494/2; 1770/
C;Superfamily: rat MEGF1 protein; cadherin repeat homology; EGF homology; laminin G repe
F;3761-3900/Domain: laminin G repeat homology <LGR>

  Query Match             15.3%;  Score 170.5;  DB 2;  Length 4307;
  Best Local Similarity   25.7%;  Pred. No. 3e-05;
```

```
  Matches   55;  Conservative   38;  Mismatches   70;  Indels   51;  Gaps   10;

Qy          28 DGFLAFPGHVFSRSLPEVPET---IELEVRTSTASGLLLWQGVEVGEAGQGK-DFISLGL 83
               ||:       |  ||    ||   :|:|::|:| :|:::|         :|| ||  | |
Db        3744 DGY-------FEISLSNEIETRMELEIELKTTTHNGIIMW--------SRGKRDFHMLRL 3788

Qy          84 QDGHLVFRYQLGSGEARLVSEDPINDGEWHRVTALREGRRGSIQVDGEELVSGRSPGPNV 143
                :|   : :  |:|   : |:  : ||:|||:   |  ||  : || |:|    ||  :
Db        3789 VNGTPEYHWNAGTGTGIVTSKTSVVDGQWHRIAISRRQRRTRMTVDDEDLQEAFSPIGST 3848

Qy         144 AVNA---KGSVYIGGAPDVATLTGGRFSSGITGCVK--------------NLVLHSARPG 186
                :|       : :|     | :  |  : |:: | :               : |  |:||
Db        3849 VINLHRYSQKLVLG-----AKVDDGELTDGVSACFRTISVDGMKVLKTRQGMKLFGAQPG 3903

Qy         187 -----APPPQPLDLQHR----AQAGANTR-PCPS 210
                    : |   |  ||     :|  :: :  |||
Db        3904 CSALTSSPCNDLPCQHAGTCISQGKSHFKCECPS 3937


RESULT 14
A48216
neurexin III-alpha secreted type 1 precursor - rat
C;Species: Rattus norvegicus (Norway rat)
C;Date: 26-May-1994 #sequence_revision 26-May-1994 #text_change 09-Dec-2002
C;Accession: A48216; B48216
R;Ushkaryov, Y.A.; Suedhof, T.C.
Proc. Natl. Acad. Sci. U.S.A. 90, 6410-6414, 1993
A;Title: Neurexin IIIalpha: extensive alternative splicing generates membrane-bound and
A;Reference number: A48216; MUID:93342001; PMID:8341647
A;Accession: A48216
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-1438 <USH>
A;Cross-references: GB:L14851
A;Accession: B48216
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-1368,1372-1438 <US2>
A;Cross-references: GB:L14851
C;Genetics:
A;Introns: 1372/1
C;Superfamily: neurexin; EGF homology
C;Keywords: alternative splicing; brain; cell surface component; duplication; extracellu
F;1-27/Domain: signal sequence #status predicted <SIG>
F;202-234/Domain: EGF homology <EGF>
F;651-683/Domain: EGF homology <EGF1>

  Query Match             15.1%;  Score 168;  DB 2;  Length 1438;
  Best Local Similarity   30.7%;  Pred. No. 1.3e-05;
  Matches   42;  Conservative   23;  Mismatches   60;  Indels   12;  Gaps    3;

Qy          47 ETIELEVRTSTASGLLLWQGVEVGEAGQGKDFISLGLQDGHLVFRYQLGSGEARLVSEDP 106
               : | |  :|   :||:|         |:  |:::| |:|| :     ||||    : | |
Db         285 DEITLSFKTWQRNGLIL-------HTGKSADYVNLALKDGAVSLVINLGSGAFEAIVE-P 336

Qy         107 I----NDGEWHRVTALREGRRGSIQVDGEELVSGRSPGPNVAVNAKGSVYIGGAPDVATL 162
               :    ||  || |   |  |: :| |||    :| :      : :    |:||:|  | |
Db         337 VNGKFNDNAWHDVKVTRNLRQVTISVDGILTTTGYTQEDYTMLGSDDFFYVGGSPSTADL 396

Qy         163 TGGRFSSGITGCVKNLV 179
                |   |:   ||:| :|
Db         397 PGSPVSNNFMGCLKEVV 413


RESULT 15
B48218
neurexin III-alpha membrane-bound type 3 precursor - rat
C;Species: Rattus norvegicus (Norway rat)
C;Date: 26-May-1994 #sequence_revision 26-May-1994 #text_change 09-Dec-2002
C;Accession: B48218; C48218
R;Ushkaryov, Y.A.; Suedhof, T.C.
Proc. Natl. Acad. Sci. U.S.A. 90, 6410-6414, 1993
A;Title: Neurexin IIIalpha: extensive alternative splicing generates membrane-bound and
```

```
A;Reference number: A48216; MUID:93342001; PMID:8341647
A;Accession: B48218
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-1471 <USH>
A;Cross-references: GB:L14851
A;Accession: C48218
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-1368,1372-1471 <US2>
A;Cross-references: GB:L14851
C;Genetics:
A;Introns: 1372/1
C;Superfamily: neurexin; EGF homology
C;Keywords: alternative splicing; brain; cell surface component; duplication; receptor;
F;1-27/Domain: signal sequence #status predicted <SIG>
F;202-234/Domain: EGF homology <EGF>
F;651-683/Domain: EGF homology <EGF1>

  Query Match             15.1%;  Score 168;  DB 2;  Length 1471;
  Best Local Similarity   30.7%;  Pred. No. 1.3e-05;
  Matches   42;  Conservative   23;  Mismatches   60;  Indels   12;  Gaps    3;

Qy          47 ETIELEVRTSTASGLLLWQGVEVGEAGQGKDFISLGLQDGHLVFRYQLGSGEARLVSEDP 106
               : | |  :|   :||:|         |:  |:::| |:|| :     ||||    : | |
Db         285 DEITLSFKTWQRNGLIL-------HTGKSADYVNLALKDGAVSLVINLGSGAFEAIVE-P 336

Qy         107 I----NDGEWHRVTALREGRRGSIQVDGEELVSGRSPGPNVAVNAKGSVYIGGAPDVATL 162
               :    ||  || |   |  |: :| |||    :| :      : :    |:||:|  | |
Db         337 VNGKFNDNAWHDVKVTRNLRQVTISVDGILTTTGYTQEDYTMLGSDDFFYVGGSPSTADL 396

Qy         163 TGGRFSSGITGCVKNLV 179
                |   |:   ||:| :|
Db         397 PGSPVSNNFMGCLKEVV 413


Search completed: March  9, 2004, 17:22:36
Job time : 8.18098 secs
```

```
                    GenCore version 5.1.6
          Copyright (c) 1993 - 2004  Compugen Ltd.

OM protein - protein search, using sw model

Run on:         March  9, 2004, 17:11:52 ; Search time 5.07923 Seconds
                                           (without alignments)
                                           2152.832 Million cell updates/sec

Title:          US-10-006-011A-10
Perfect score:  1115
Sequence:       1 GIAESDWHLEGSGGNDAPGQ..........QPLDLQHRAQAGANTRPCPS 210

Scoring table:  BLOSUM62
                Gapop 10.0 , Gapext 0.5

Searched:       141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters:      141681

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :       SwissProt_42:*

      Pred. No. is the number of results predicted by chance to have a
      score greater than or equal to the score of the result being printed,
      and is derived by analysis of the total score distribution.

                                SUMMARIES
                        %
Result               Query
  No.     Score      Match Length DB  ID                         Description
-------------------------------------------------------------------------------
     1     1115     100.0   4391  1  PGBM_HUMAN                 P98160 homo sapien
     2    964.5      86.5   3707  1  PGBM_MOUSE                 Q05793 mus musculu
     3    338.5      30.4   3375  1  UN52_CAEEL                 Q06561 caenorhabdi
     4    242.5      21.7   1959  1  AGRI_RAT                   P25304 rattus norv
     5      235      21.1   1328  1  AGRI_DISOM                 Q90404 discopyge o
     6    223.5      20.0   3110  1  LMA2_HUMAN                 P24043 homo sapien
     7    207.5      18.6   3106  1  LMA2_MOUSE                 Q60675 mus musculu
     8    200.5      18.0   1955  1  AGRI_CHICK                 P31696 gallus gall
     9      191      17.1   1712  1  NX2A_HUMAN                 Q9p2s2 homo sapien
    10    188.5      16.9   3075  1  LMA1_HUMAN                 P25391 homo sapien
    11      188      16.9   1715  1  NX2A_RAT                   Q63374 rattus norv
    12    185.5      16.6   3672  1  LML2_CAEEL                 Q21313 caenorhabdi
    13    173.5      15.6   1816  1  LMA4_MOUSE                 P97927 mus musculu
    14      171      15.3   3333  1  LMA3_MOUSE                 Q61789 mus musculu
    15      168      15.1   1541  1  NX3A_HUMAN                 Q9y4c0 homo sapien
    16    165.5      14.8   1816  1  LMA4_HUMAN                 Q16363 homo sapien
    17      165      14.8   1578  1  NX3A_RAT                   Q07310 rattus norv
    18    162.5      14.6   1713  1  LMA3_HUMAN                 Q16787 homo sapien
    19      162      14.5   3084  1  LMA1_MOUSE                 P19137 mus musculu
    20    160.5      14.4   3712  1  LMA_DROME                  Q00174 drosophila
    21      159      14.3   1363  1  NX1A_CHICK                 Q9ddd0 gallus gall
    22      159      14.3   1514  1  NX1A_RAT                   Q63372 rattus norv
    23      159      14.3   1530  1  NX1A_BOVIN                 Q28146 bos taurus
    24    156.5      14.0   1331  1  CTA2_HUMAN                 Q9uhc6 homo sapien
    25      156      14.0   1477  1  NX1A_HUMAN                 Q9ulb1 homo sapien
    26    152.5      13.7   4705  1  FAT2_DROME                 Q9vw71 drosophila
    27      151      13.5   1308  1  CTA4_HUMAN                 Q9c0a0 homo sapien
    28      151      13.5   1529  1  SLT2_HUMAN                 O94813 homo sapien
    29      147      13.2   1288  1  CTA3_HUMAN                 Q9bz76 homo sapien
    30    139.5      12.5   5147  1  FAT_DROME                  P33450 drosophila
    31    138.5      12.4   3695  1  LMA5_HUMAN                 O15230 homo sapien
    32      137      12.3   1284  1  NRX4_DROME                 Q94887 drosophila
    33    130.5      11.7   1381  1  CTA1_RAT                   P97846 rattus norv
    34      130      11.7   4590  1  FATH_HUMAN                 Q14517 homo sapien
    35    129.5      11.6   1384  1  CTA1_HUMAN                 P78357 homo sapien
    36      129      11.6   1310  1  CTA4_MOUSE                 Q99p47 mus musculu
    37    125.5      11.3   3097  1  CADN_DROME                 O15943 drosophila
    38      125      11.2   2923  1  CLR2_HUMAN                 Q9hcu4 homo sapien
    39    123.5      11.1   1385  1  CTA1_MOUSE                 O54991 mus musculu
    40      122      10.9   2144  1  CLR2_RAT                   Q9qyp2 rattus norv
    41    121.5      10.9    529  1  NX1A_MOUSE                 Q9cs84 mus musculu
    42    121.5      10.9   3579  1  STAN_DROME                 Q9v5n8 drosophila
    43    120.5      10.8   2215  1  CDN2_DROME                 Q9vjb6 drosophila
    44    118.5      10.6   2920  1  CLR2_MOUSE                 Q9r0m0 mus musculu
    45    116.5      10.4   3718  1  LMA5_MOUSE                 Q61001 mus musculu

                                ALIGNMENTS

RESULT 1
PGBM_HUMAN
ID   PGBM_HUMAN     STANDARD;      PRT;  4391 AA.
AC   P98160; Q16287; Q9H3V5;
DT   01-OCT-1996 (Rel. 34, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Basement membrane-specific heparan sulfate proteoglycan core
DE   protein precursor (HSPG) (Perlecan) (PLC).
GN   HSPG2.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=92112994; PubMed=1730768;
RA   Kallunki P., Tryggvason K.;
RT   "Human basement membrane heparan sulfate proteoglycan core protein: a
RT   467-kD protein containing multiple domains resembling elements of the
RT   low density lipoprotein receptor, laminin, neural cell adhesion
RT   molecules, and epidermal growth factor.";
RL   J. Cell Biol. 116:559-571(1992).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Colon, and Skin;
RX   MEDLINE=92235084; PubMed=1569102;
RA   Murdoch A.D., Dodge G.R., Cohen I., Tuan R.S., Iozzo R.V.;
RT   "Primary structure of the human heparan sulfate proteoglycan from
RT   basement membrane (HSPG2/perlecan). A chimeric molecule with multiple
RT   domains homologous to the low density lipoprotein receptor, laminin,
RT   neural cell adhesion molecules, and epidermal growth factor.";
RL   J. Biol. Chem. 267:8544-8557(1992).
RN   [3]
RP   SEQUENCE OF 22-4391 FROM N.A., AND VARIANT SJS1 TYR-1532.
RX   MEDLINE=20553141; PubMed=11101850;
RA   Nicole S., Davoine C.-S., Topaloglu H., Cattolico L., Barral D.,
RA   Beighton P., Ben-Hamida C., Hammouda H., Cruaud C., White P.S.,
RA   Samson D., Urtizberea J.A., Lehmann-Horn F., Weissenbach J.,
RA   Hentati F., Fontaine B.;
RT   "Perlecan, the major proteoglycan of basement membranes, is altered in
RT   patients with Schwartz-Jampel syndrome (chondrodystrophic myotonia).";
RL   Nat. Genet. 26:480-483(2000).
RN   [4]
RP   SEQUENCE OF 1016-1470 FROM N.A.
RC   TISSUE=Colon;
RX   MEDLINE=91365376; PubMed=1679749;
RA   Dodge G.R., Kovalszky I., Chu M.L., Hassell J.R., McBride O.W.,
RA   Yi H.F., Iozzo R.V.;
RT   "Heparan sulfate proteoglycan of human colon: partial molecular
RT   cloning, cellular expression, and mapping of the gene (HSPG2) to the
RT   short arm of human chromosome 1.";
RL   Genomics 10:673-680(1991).
RN   [5]
RP   SEQUENCE OF 890-1396 FROM N.A.
RC   TISSUE=Fibrosarcoma;
```

```
RX   MEDLINE=92120660; PubMed=1685141;
RA   Kallunki P., Eddy R.L., Byers M.G., Kestila M., Shows T.B.,
RA   Tryggvason K.;
RT   "Cloning of human heparan sulfate proteoglycan core protein,
RT   assignment of the gene (HSPG2) to 1p36.1-->p35 and identification of
RT   a BamHI restriction fragment length polymorphism.";
RL   Genomics 11:389-396(1991).
RN   [6]
RP   SEQUENCE OF 1-21 FROM N.A.
RX   MEDLINE=94052171; PubMed=8234307;
RA   Cohen I.R., Graessel S., Murdoch A.D., Iozzo R.V.;
RT   "Structural characterization of the complete human perlecan gene and
RT   its promoter.";
RL   Proc. Natl. Acad. Sci. U.S.A. 90:10404-10408(1993).
RN   [7]
RP   CARBOHYDRATE-LINKAGE SITE ASN-2121.
RX   MEDLINE=22660472; PubMed=12754519;
RA   Zhang H., Li X.-J., Martin D.B., Aebersold R.;
RT   "Identification and quantification of N-linked glycoproteins using
RT   hydrazide chemistry, stable isotope labeling and mass spectrometry.";
RL   Nat. Biotechnol. 21:660-666(2003).
CC   -!- FUNCTION: This protein is an integral component of basement
CC       membranes. It is responsible for the fixed negative electrostatic
CC       charge and is involved in the charge-selective ultrafiltration
CC       properties. It serves as an attachment substrate for cells.
CC   -!- SUBUNIT: Purified perlecan has a strong tendency to aggregate in
CC       dimers or stellate structures. It interacts with other basement
CC       membrane components such as laminin, prolargin and collagen type
CC       IV.
CC   -!- SUBCELLULAR LOCATION: Extracellular.
CC   -!- TISSUE SPECIFICITY: Found in the basement membranes.
CC   -!- PTM: CONTAINS THREE HEPARAN SULFATE CHAINS AS WELL AS N-LINKED
CC       AND O-LINKED OLIGOSACCHARIDES.
CC   -!- DISEASE: Defects in HSPG2 are the cause of Schwartz-Jampel
CC       syndrome (SJS1) [MIM:255800]; a rare autosomal recessive disorder
CC       characterized by permanent myotonia (prolonged failure of muscle
CC       relaxation) and skeletal dysplasia, resulting in reduced stature,
CC       kyphoscoliosis, bowing of the diaphyses and irregular epiphyses.
CC   -!- SIMILARITY: Contains 4 LDL-receptor class A domains.
CC   -!- SIMILARITY: Contains 11 laminin EGF-like domains.
CC   -!- SIMILARITY: Contains 3 laminin IV domains.
CC   -!- SIMILARITY: Contains 22 immunoglobulin-like C2-type domains.
CC   -!- SIMILARITY: Contains 3 laminin G-like domains.
CC   -!- SIMILARITY: Contains 4 EGF-like domains.
CC   -!- SIMILARITY: Contains 1 SEA domain.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; X62515; CAA44373.1; -.
DR   EMBL; M85289; AAA52700.1; -.
DR   EMBL; AL445795; CAC18534.1; -.
DR   EMBL; M64283; AAA52699.1; -.
DR   EMBL; S76436; AAB21121.2; -.
DR   EMBL; L22078; -; NOT_ANNOTATED_CDS.
DR   PIR; A38096; A38096.
DR   HSSP; P00740; 1EDM.
DR   Siena-2DPAGE; P98160; -.
DR   Genew; HGNC:5273; HSPG2.
DR   MIM; 142461; -.
DR   MIM; 255800; -.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR000742; EGF_2.
DR   InterPro; IPR006209; EGF_like.
DR   InterPro; IPR006210; IEGF.
DR   InterPro; IPR007110; Ig-like.
DR   InterPro; IPR003599; Ig.
DR   InterPro; IPR003598; Ig_c2.
DR   InterPro; IPR003596; Ig_v.
DR   InterPro; IPR000034; Laminin_B.
DR   InterPro; IPR002049; Laminin_EGF.
DR   InterPro; IPR001791; Laminin_G.
DR   InterPro; IPR002172; LDL_receptor_A.
DR   InterPro; IPR000082; SEA_domain.
DR   Pfam; PF00008; EGF; 4.
DR   Pfam; PF00047; ig; 22.
DR   Pfam; PF00052; laminin_B; 3.
DR   Pfam; PF00053; laminin_EGF; 7.
DR   Pfam; PF00054; laminin_G; 3.
DR   Pfam; PF00057; ldl_recept_a; 4.
DR   Pfam; PF01390; SEA; 1.
DR   PRINTS; PR00261; LDLRECEPTOR.
DR   ProDom; PD003031; Laminin_B; 3.
DR   SMART; SM00181; EGF; 15.
DR   SMART; SM00180; EGF_Lam; 12.
DR   SMART; SM00409; IG; 22.
DR   SMART; SM00408; IGc2; 21.
DR   SMART; SM00406; IGv; 7.
DR   SMART; SM00281; LamB; 3.
DR   SMART; SM00282; LamG; 3.
DR   SMART; SM00192; LDLa; 4.
DR   SMART; SM00200; SEA; 1.
DR   PROSITE; PS00022; EGF_1; 9.
DR   PROSITE; PS01186; EGF_2; 6.
DR   PROSITE; PS50026; EGF_3; 4.
DR   PROSITE; PS50835; IG_LIKE; 22.
DR   PROSITE; PS50025; LAM_G_DOMAIN; 3.
DR   PROSITE; PS01248; LAMININ_TYPE_EGF; 11.
DR   PROSITE; PS01209; LDLRA_1; 4.
DR   PROSITE; PS50068; LDLRA_2; 4.
DR   PROSITE; PS50024; SEA; 1.
KW   Signal; Basement membrane; Proteoglycan; Repeat; Glycoprotein;
KW   Heparan sulfate; Laminin EGF-like domain; Immunoglobulin domain;
KW   Extracellular matrix; EGF-like domain; Disease mutation.
FT   SIGNAL        1     21       POTENTIAL.
FT   CHAIN        22   4391       BASEMENT MEMBRANE-SPECIFIC HEPARAN
FT                                SULFATE PROTEOGLYCAN CORE PROTEIN.
FT   DOMAIN       80    194       SEA.
FT   DOMAIN      198    235       LDL-RECEPTOR CLASS A 1.
FT   DOMAIN      284    320       LDL-RECEPTOR CLASS A 2.
FT   DOMAIN      324    360       LDL-RECEPTOR CLASS A 3.
FT   DOMAIN      367    404       LDL-RECEPTOR CLASS A 4.
FT   DOMAIN      405    504       IG-LIKE C2-TYPE 1.
FT   DOMAIN      521    530       LAMININ EGF-LIKE 1 (N-TERMINAL).
FT   DOMAIN      531    730       LAMININ DOMAIN IV 1 (DOMAIN III A).
FT   DOMAIN      731    763       LAMININ EGF-LIKE 1 (C-TERMINAL).
FT   DOMAIN      764    813       LAMININ EGF-LIKE 2.
FT   DOMAIN      814    871       LAMININ EGF-LIKE 3.
FT   DOMAIN      879    923       LAMININ EGF-LIKE 4 (INCOMPLETE).
FT   DOMAIN      924    933       LAMININ EGF-LIKE 5 (N-TERMINAL).
FT   DOMAIN      934   1125       LAMININ DOMAIN IV 2 (DOMAIN III B).
FT   DOMAIN     1126   1158       LAMININ EGF-LIKE 5 (C-TERMINAL).
FT   DOMAIN     1159   1208       LAMININ EGF-LIKE 6.
FT   DOMAIN     1209   1265       LAMININ EGF-LIKE 7.
FT   DOMAIN     1275   1324       LAMININ EGF-LIKE 8.
FT   DOMAIN     1325   1334       LAMININ EGF-LIKE 9 (N-TERMINAL).
FT   DOMAIN     1335   1529       LAMININ DOMAIN IV 3 (DOMAIN III C).
FT   DOMAIN     1530   1562       LAMININ EGF-LIKE 9 (C-TERMINAL).
FT   DOMAIN     1563   1612       LAMININ EGF-LIKE 10.
FT   DOMAIN     1613   1670       LAMININ EGF-LIKE 11.
FT   DOMAIN     1677   1771       IG-LIKE C2-TYPE 2.
FT   DOMAIN     1772   1865       IG-LIKE C2-TYPE 3.
FT   DOMAIN     1866   1955       IG-LIKE C2-TYPE 4.
FT   DOMAIN     1956   2051       IG-LIKE C2-TYPE 5.
FT   DOMAIN     2052   2151       IG-LIKE C2-TYPE 6.
FT   DOMAIN     2152   2244       IG-LIKE C2-TYPE 7.
FT   DOMAIN     2245   2340       IG-LIKE C2-TYPE 8.
FT   DOMAIN     2341   2436       IG-LIKE C2-TYPE 9.
FT   DOMAIN     2437   2533       IG-LIKE C2-TYPE 10.
FT   DOMAIN     2534   2629       IG-LIKE C2-TYPE 11.
FT   DOMAIN     2630   2726       IG-LIKE C2-TYPE 12.
```

```
  Query Match             100.0%;  Score 1115;  DB 1;  Length 4391;
  Best Local Similarity   100.0%;  Pred. No. 2e-82;
  Matches  210;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 GIAESDWHLEGSGGNDAPGQYGAYFHDDGFLAFPGHVFSRSLPEVPETIELEVRTSTASG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4182 GIAESDWHLEGSGGNDAPGQYGAYFHDDGFLAFPGHVFSRSLPEVPETIELEVRTSTASG 4241

Qy         61 LLLWQGVEVGEAGQGKDFISLGLQDGHLVFRYQLGSGEARLVSEDPINDGEWHRVTALRE 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4242 LLLWQGVEVGEAGQGKDFISLGLQDGHLVFRYQLGSGEARLVSEDPINDGEWHRVTALRE 4301

Qy        121 GRRGSIQVDGEELVSGRSPGPNVAVNAKGSVYIGGAPDVATLTGGRFSSGITGCVKNLVL 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4302 GRRGSIQVDGEELVSGRSPGPNVAVNAKGSVYIGGAPDVATLTGGRFSSGITGCVKNLVL 4361

Qy        181 HSARPGAPPPQPLDLQHRAQAGANTRPCPS 210
              ||||||||||||||||||||||||||||||
Db       4362 HSARPGAPPPQPLDLQHRAQAGANTRPCPS 4391

RESULT 2
PGBM_MOUSE
ID   PGBM_MOUSE     STANDARD;      PRT;  3707 AA.
AC   Q05793;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Basement membrane-specific heparan sulfate proteoglycan core
DE   protein precursor (HSPG) (Perlecan) (PLC).
GN   HSPG2.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Melanoma;
RX   MEDLINE=92078153; PubMed=1744087;
RA   Noonan D.M., Fulle A., Valente P., Cai S., Horigan E., Sasaki M.,
RA   Yamada Y., Hassell J.R.;
RT   "The complete sequence of perlecan, a basement membrane heparan
RT   sulfate proteoglycan, reveals extensive similarity with laminin A
RT   chain, low density lipoprotein-receptor, and the neural cell adhesion
RT   molecule.";
RL   J. Biol. Chem. 266:22939-22947(1991).
RN   [2]
RP   SEQUENCE OF 940-1601 AND 1870-2600 FROM N.A., AND PARTIAL SEQUENCE.
RX   MEDLINE=89034110; PubMed=2972708;
RA   Noonan D.M., Horigan E.A., Ledbetter S.R., Vogeli G., Sasaki M.,
RA   Yamada Y., Hassell J.R.;
RT   "Identification of cDNA clones encoding different domains of the
RT   basement membrane heparan sulfate proteoglycan.";
RL   J. Biol. Chem. 263:16379-16387(1988).
CC   -!- FUNCTION: This protein is an integral component of basement
CC       membranes. It is responsible for the fixed negative electrostatic
CC       charge and is involved in the charge-selective ultrafiltration
CC       properties. It serves as an attachment substrate for cells.
CC   -!- SUBUNIT: Purified perlecan has a strong tendency to aggregate in
CC       dimers or stellate structures. It interacts with other basement
CC       membrane components such as laminin, prolargin and collagen type
CC       IV.
CC   -!- SUBCELLULAR LOCATION: Extracellular.
CC   -!- TISSUE SPECIFICITY: Found in the basement membranes.
CC   -!- PTM: CONTAINS THREE HEPARAN SULFATE CHAINS AS WELL AS N-LINKED
CC       AND O-LINKED OLIGOSACCHARIDES.
CC   -!- SIMILARITY: Contains 4 LDL-receptor class A domains.
CC   -!- SIMILARITY: Contains 11 laminin EGF-like domains.
CC   -!- SIMILARITY: Contains 3 laminin IV domains.
CC   -!- SIMILARITY: Contains 15 immunoglobulin-like C2-type domains.
CC   -!- SIMILARITY: Contains 3 laminin G-like domains.
CC   -!- SIMILARITY: Contains 1 EGF-like domain.
CC   -!- SIMILARITY: Contains 1 SEA domain.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; M77174; AAA39911.1; -.
DR   EMBL; J04054; AAA39899.1; -.
DR   EMBL; J04055; AAA39912.1; -.
DR   PIR; S18252; S18252.
DR   PDB; 1GL4; 28-NOV-01.
DR   MGD; MGI:96257; Hspg2.
DR   GO; GO:0005604; C:basement membrane; IDA.
DR   GO; GO:0008104; P:protein localization; IMP.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR000742; EGF_2.
DR   InterPro; IPR006209; EGF_like.
DR   InterPro; IPR007110; Ig-like.
DR   InterPro; IPR003598; Ig_c2.
DR   InterPro; IPR000034; Laminin_B.
DR   InterPro; IPR002049; Laminin_EGF.
DR   InterPro; IPR001791; Laminin_G.
DR   InterPro; IPR002172; LDL_receptor_A.
DR   InterPro; IPR000082; SEA_domain.
DR   Pfam; PF00008; EGF; 4.
DR   Pfam; PF00047; ig; 15.
DR   Pfam; PF00052; laminin_B; 3.
DR   Pfam; PF00053; laminin_EGF; 7.
DR   Pfam; PF00054; laminin_G; 3.
DR   Pfam; PF00057; ldl_recept_a; 4.
DR   Pfam; PF01390; SEA; 1.
DR   PRINTS; PR00261; LDLRECEPTOR.
DR   ProDom; PD003031; Laminin_B; 3.
DR   SMART; SM00180; EGF_Lam; 7.
DR   SMART; SM00408; IGc2; 14.
DR   SMART; SM00281; LamB; 3.
DR   SMART; SM00282; LamG; 3.
DR   SMART; SM00192; LDLa; 4.
DR   SMART; SM00200; SEA; 1.
DR   PROSITE; PS00022; EGF_1; 8.
DR   PROSITE; PS01186; EGF_2; 5.
DR   PROSITE; PS50026; EGF_3; 4.
DR   PROSITE; PS50835; IG_LIKE; 15.
DR   PROSITE; PS50025; LAM_G_DOMAIN; 3.
DR   PROSITE; PS01248; LAMININ_TYPE_EGF; 11.
DR   PROSITE; PS01209; LDLRA_1; 4.
DR   PROSITE; PS50068; LDLRA_2; 4.
DR   PROSITE; PS50024; SEA; 1.
KW   Signal; Basement membrane; Proteoglycan; Repeat; Glycoprotein;
KW   Heparan sulfate; Laminin EGF-like domain; Immunoglobulin domain;
KW   Extracellular matrix; EGF-like domain; 3D-structure.
FT   SIGNAL        1     21       POTENTIAL.
FT   CHAIN        22   3707       BASEMENT MEMBRANE-SPECIFIC HEPARAN
FT                                SULFATE PROTEOGLYCAN CORE PROTEIN.
FT   DOMAIN       80    194       SEA.
FT   DOMAIN      195    234       LDL-RECEPTOR CLASS A 1.
FT   DOMAIN      281    319       LDL-RECEPTOR CLASS A 2.
FT   DOMAIN      320    359       LDL-RECEPTOR CLASS A 3.
FT   DOMAIN      360    403       LDL-RECEPTOR CLASS A 4.
FT   DOMAIN      404    504       IG-LIKE C2-TYPE 1.
FT   DOMAIN      521    530       LAMININ EGF-LIKE 1 (N-TERMINAL).
FT   DOMAIN      531    730       LAMININ DOMAIN IV 1 (DOMAIN III A).
FT   DOMAIN      731    763       LAMININ EGF-LIKE 1 (C-TERMINAL).
FT   DOMAIN      764    813       LAMININ EGF-LIKE 2.
FT   DOMAIN      814    871       LAMININ EGF-LIKE 3.
FT   DOMAIN      879    923       LAMININ EGF-LIKE 4 (INCOMPLETE).
FT   DOMAIN      924    933       LAMININ EGF-LIKE 5 (N-TERMINAL).
FT   DOMAIN      934   1125       LAMININ DOMAIN IV 2 (DOMAIN III B).
FT   DOMAIN     1126   1158       LAMININ EGF-LIKE 5 (C-TERMINAL).
```

```
FT   DOMAIN     1159   1208       LAMININ EGF-LIKE 6.
FT   DOMAIN     1209   1265       LAMININ EGF-LIKE 7.
FT   DOMAIN     1275   1324       LAMININ EGF-LIKE 8.
FT   DOMAIN     1325   1334       LAMININ EGF-LIKE 9 (N-TERMINAL).
FT   DOMAIN     1335   1529       LAMININ DOMAIN IV 3 (DOMAIN III C).
FT   DOMAIN     1530   1562       LAMININ EGF-LIKE 9 (C-TERMINAL).
FT   DOMAIN     1563   1612       LAMININ EGF-LIKE 10.
FT   DOMAIN     1613   1670       LAMININ EGF-LIKE 11.
FT   DOMAIN     1677   1771       IG-LIKE C2-TYPE 2.
FT   DOMAIN     1772   1865       IG-LIKE C2-TYPE 3.
FT   DOMAIN     1866   1954       IG-LIKE C2-TYPE 4.
FT   DOMAIN     1955   2049       IG-LIKE C2-TYPE 5.
FT   DOMAIN     2050   2148       IG-LIKE C2-TYPE 6.
FT   DOMAIN     2149   2244       IG-LIKE C2-TYPE 7.
FT   DOMAIN     2245   2343       IG-LIKE C2-TYPE 8.
FT   DOMAIN     2344   2436       IG-LIKE C2-TYPE 9.
FT   DOMAIN     2437   2532       IG-LIKE C2-TYPE 10.
FT   DOMAIN     2533   2619       IG-LIKE C2-TYPE 11.
FT   DOMAIN     2620   2720       IG-LIKE C2-TYPE 12.
FT   DOMAIN     2721   2809       IG-LIKE C2-TYPE 13.
FT   DOMAIN     2810   2895       IG-LIKE C2-TYPE 14.
FT   DOMAIN     2896   2980       IG-LIKE C2-TYPE 15.
FT   DOMAIN     2984   3162       LAMININ G-LIKE 1.
FT   DOMAIN     3163   3241       EGF-LIKE.
FT   DOMAIN     3245   3425       LAMININ G-LIKE 2.
FT   DOMAIN     3518   3705       LAMININ G-LIKE 3.
FT   SITE         65     67       HEPARAN SULFATE (POTENTIAL).
FT   SITE         71     73       HEPARAN SULFATE (POTENTIAL).
FT   SITE         76     78       HEPARAN SULFATE (POTENTIAL).
FT   SITE       3615   3617       MEDIATES MOTOR NEURON ATTACHMENT
FT                                (POTENTIAL).
FT   DISULFID    199    212       BY SIMILARITY.
FT   DISULFID    206    225       BY SIMILARITY.
FT   DISULFID    219    234       BY SIMILARITY.
FT   DISULFID    285    297       BY SIMILARITY.
FT   DISULFID    292    310       BY SIMILARITY.
FT   DISULFID    304    319       BY SIMILARITY.
FT   DISULFID    325    337       BY SIMILARITY.
FT   DISULFID    332    350       BY SIMILARITY.
FT   DISULFID    344    359       BY SIMILARITY.
FT   DISULFID    368    381       BY SIMILARITY.
FT   DISULFID    375    394       BY SIMILARITY.
FT   DISULFID    388    403       BY SIMILARITY.
FT   DISULFID    428    479       BY SIMILARITY.
FT   DISULFID    764    773       BY SIMILARITY.
FT   DISULFID    766    780       BY SIMILARITY.
FT   DISULFID    783    792       BY SIMILARITY.
FT   DISULFID    795    811       BY SIMILARITY.
FT   DISULFID    814    829       BY SIMILARITY.
FT   DISULFID    816    839       BY SIMILARITY.
FT   DISULFID    842    851       BY SIMILARITY.
FT   DISULFID    854    869       BY SIMILARITY.
FT   DISULFID   1159   1168       BY SIMILARITY.
FT   DISULFID   1161   1175       BY SIMILARITY.
FT   DISULFID   1178   1187       BY SIMILARITY.
FT   DISULFID   1190   1206       BY SIMILARITY.
FT   DISULFID   1209   1224       BY SIMILARITY.
FT   DISULFID   1211   1234       BY SIMILARITY.
FT   DISULFID   1237   1246       BY SIMILARITY.
FT   DISULFID   1249   1263       BY SIMILARITY.
FT   DISULFID   1275   1287       BY SIMILARITY.
FT   DISULFID   1277   1293       BY SIMILARITY.
FT   DISULFID   1295   1304       BY SIMILARITY.
FT   DISULFID   1307   1322       BY SIMILARITY.
FT   DISULFID   1563   1572       BY SIMILARITY.
FT   DISULFID   1565   1579       BY SIMILARITY.
FT   DISULFID   1582   1591       BY SIMILARITY.
FT   DISULFID   1594   1610       BY SIMILARITY.
FT   DISULFID   1613   1628       BY SIMILARITY.
FT   DISULFID   1615   1638       BY SIMILARITY.
FT   DISULFID   1641   1650       BY SIMILARITY.
FT   DISULFID   1653   1668       BY SIMILARITY.
FT   DISULFID   1792   1839       BY SIMILARITY.
```

```
FT   DISULFID   1886   1932       BY SIMILARITY.
FT   DISULFID   1976   2021       BY SIMILARITY.
FT   DISULFID   2073   2118       BY SIMILARITY.
FT   DISULFID   2170   2215       BY SIMILARITY.
FT   DISULFID   2268   2313       BY SIMILARITY.

  Query Match             86.5%;  Score 964.5;  DB 1;  Length 3707;
  Best Local Similarity   85.7%;  Pred. No. 3e-70;
  Matches  180;  Conservative   13;  Mismatches   16;  Indels    1;  Gaps    1;

Qy          1 GIAESDWHLEGSGGNDAPGQYGAYFHDDGFLAFPGHVFSRSLPEVPETIELEVRTSTASG 60
              |: ||||| ||||||||||||||||:|:|||  ||: ||||||||||||| ||||||| |
Db       3499 GVVESDWHPEGSGGNDAPGQYGAYFYDNGFLGLPGNSFSRSLPEVPETIEFEVRTSTADG 3558

Qy         61 LLLWQGVEVGEAGQGKDFISLGLQDGHLVFRYQLGSGEARLVSEDPINDGEWHRVTALRE 120
              ||||||| | || : ||||||||||||||| |||||||||||| ||||||||||:|||||
Db       3559 LLLWQGV-VREASRSKDFISLGLQDGHLVFSYQLGSGEARLVSGDPINDGEWHRITALRE 3617

Qy        121 GRRGSIQVDGEELVSGRSPGPNVAVNAKGSVYIGGAPDVATLTGGRFSSGITGCVKNLVL 180
              |:|||||||||:||:||||||||||| |  :|||||||||||| |:||||||||:|||||
Db       3618 GQRGSIQVDGEDLVTGRSPGPNVAVNTKDIIYIGGAPDVATLTRGKFSSGITGCIKNLVL 3677

Qy        181 HSARPGAPPPQPLDLQHRAQAGANTRPCPS 210
              |:||||||||||||||||||||||||||||
Db       3678 HTARPGAPPPQPLDLQHRAQAGANTRPCPS 3707

RESULT 3
UN52_CAEEL
ID   UN52_CAEEL     STANDARD;      PRT;  3375 AA.
AC   Q06561; O18261; O18263; Q9XTD2; Q9XTI5;
DT   01-JUN-1994 (Rel. 29, Created)
DT   10-OCT-2003 (Rel. 42, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Basement membrane proteoglycan precursor (Perlecan homolog)
DE   (Uncoordinated protein 52).
GN   UNC-52 OR ZC101.2.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC   Rhabditidae; Peloderinae; Caenorhabditis.
OX   NCBI_TaxID=6239;
RN   [1]
RP   SEQUENCE FROM N.A. (ISOFORM A), AND FUNCTION.
RX   MEDLINE=93339574; PubMed=8393416;
RA   Rogalski T.M., Williams B.D., Mullen G.P., Moerman D.G.;
RT   "Products of the unc-52 gene in Caenorhabditis elegans are homologous
RT   to the core protein of the mammalian basement membrane heparan
RT   sulfate proteoglycan.";
RL   Genes Dev. 7:1471-1484(1993).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Bristol N2;
RA   Percy C.M., Baynes C.;
RL   Submitted (MAR-1997) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   REVISIONS, AND ALTERNATIVE SPLICING.
RA   Durbin R.;
RL   Submitted (FEB-1998) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: Probable role in myofilament assembly and/or attachment
CC       of the myofilament lattice to the cell membrane. May be an
CC       extracellular anchor for integrin receptors in muscle.
CC   -!- SUBCELLULAR LOCATION: Extracellular matrix.
CC   -!- ALTERNATIVE PRODUCTS:
CC       Event=Alternative splicing; Named isoforms=4;
CC       Name=e;
CC         IsoId=Q06561-1; Sequence=Displayed;
CC         Note=No experimental confirmation available;
CC       Name=a;
CC         IsoId=Q06561-2; Sequence=VSP_007195, VSP_007196;
CC       Name=b;
CC         IsoId=Q06561-3; Sequence=VSP_007191, VSP_007192;
CC         Note=No experimental confirmation available;
CC       Name=c;
```

```
CC        IsoId=Q06561-4; Sequence=VSP_007193, VSP_007194, VSP_007195,
CC                                 VSP_007196;
CC        Note=No experimental confirmation available;
CC   -!- TISSUE SPECIFICITY: Found in the basement membrane of all
CC       contractile tissues. It is concentrated over muscle dense bodies
CC       and M-lines which are associated with beta-integrin.
CC   -!- DEVELOPMENTAL STAGE: Synthesized early in embryogenesis.
CC   -!- SIMILARITY: Contains 3 LDL-receptor class A domains.
CC   -!- SIMILARITY: Contains 16 immunoglobulin-like C2-type domains.
CC   -!- SIMILARITY: Contains 7 laminin EGF-like domains.
CC   -!- SIMILARITY: Contains 3 laminin G-like domains.
CC   -!- SIMILARITY: Contains 2 laminin IV domains.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; L13458; AAA28156.1; -.
DR   EMBL; Z93375; CAB07567.1; -.
DR   EMBL; Z93395; CAB07567.1; JOINED.
DR   EMBL; Z93375; CAB07568.1; -.
DR   EMBL; Z93395; CAB07568.1; JOINED.
DR   EMBL; Z93375; CAB07569.1; -.
DR   EMBL; Z93395; CAB07569.1; JOINED.
DR   EMBL; Z93395; CAB07704.1; -.
DR   EMBL; Z93395; CAB07706.1; -.
DR   EMBL; Z93375; CAB07706.1; JOINED.
DR   EMBL; Z93395; CAB07707.1; -.
DR   EMBL; Z93375; CAB07707.1; JOINED.
DR   EMBL; Z93395; CAB07708.1; -.
DR   EMBL; Z93375; CAB07708.1; JOINED.
DR   HSSP; P01130; 1LDR.
DR   WormPep; ZC101.2a; CE15028.
DR   WormPep; ZC101.2b; CE15030.
DR   WormPep; ZC101.2c; CE15034.
DR   WormPep; ZC101.2e; CE18424.
DR   GO; GO:0005578; C:extracellular matrix; IEP.
DR   GO; GO:0030239; P:myofibril assembly; IEP.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR001881; EGF_Ca.
DR   InterPro; IPR006209; EGF_like.
DR   InterPro; IPR006210; IEGF.
DR   InterPro; IPR007110; Ig-like.
DR   InterPro; IPR003599; Ig.
DR   InterPro; IPR003598; Ig_c2.
DR   InterPro; IPR000034; Laminin_B.
DR   InterPro; IPR002049; Laminin_EGF.
DR   InterPro; IPR001791; Laminin_G.
DR   InterPro; IPR002172; LDL_receptor_A.
DR   Pfam; PF00047; ig; 16.
DR   Pfam; PF00052; laminin_B; 2.
DR   Pfam; PF00053; laminin_EGF; 5.
DR   Pfam; PF00057; ldl_recept_a; 3.
DR   PRINTS; PR00261; LDLRECEPTOR.
DR   ProDom; PD003031; Laminin_B; 2.
DR   SMART; SM00181; EGF; 6.
DR   SMART; SM00179; EGF_CA; 2.
DR   SMART; SM00180; EGF_Lam; 6.
DR   SMART; SM00409; IG; 17.
DR   SMART; SM00408; IGc2; 17.
DR   SMART; SM00281; LamB; 1.
DR   SMART; SM00282; LamG; 3.
DR   SMART; SM00192; LDLa; 3.
DR   PROSITE; PS00022; EGF_1; 7.
DR   PROSITE; PS01186; EGF_2; 2.
DR   PROSITE; PS50026; EGF_3; 3.
DR   PROSITE; PS50835; IG_LIKE; 17.
DR   PROSITE; PS50025; LAM_G_DOMAIN; 3.
DR   PROSITE; PS01248; LAMININ_TYPE_EGF; 7.
```

```
DR   PROSITE; PS01209; LDLRA_1; 3.
DR   PROSITE; PS50068; LDLRA_2; 3.
KW   Proteoglycan; Glycoprotein; Signal; Immunoglobulin domain; Repeat;
KW   Basement membrane; Extracellular matrix; Alternative splicing;
KW   Laminin EGF-like domain.
FT   SIGNAL        1     22       POTENTIAL.
FT   CHAIN        23   3375       BASEMENT MEMBRANE PROTEOGLYCAN.
FT   DOMAIN       45    130       IG-LIKE C2-TYPE 1.
FT   DOMAIN      148    184       LDL-RECEPTOR CLASS A 1.
FT   DOMAIN      189    225       LDL-RECEPTOR CLASS A 2.
FT   DOMAIN      232    269       LDL-RECEPTOR CLASS A 3.
FT   DOMAIN      271    355       IG-LIKE C2-TYPE 2.
FT   DOMAIN      384    431       LAMININ EGF-LIKE 1 (INCOMPLETE).
FT   DOMAIN      432    441       LAMININ EGF-LIKE 2 (N-TERMINAL).
FT   DOMAIN      442    633       LAMININ DOMAIN IV 1.
FT   DOMAIN      634    666       LAMININ EGF-LIKE 2 (C-TERMINAL).
FT   DOMAIN      674    720       LAMININ EGF-LIKE 3 (INCOMPLETE).
FT   DOMAIN      721    730       LAMININ EGF-LIKE 4 (N-TERMINAL).
FT   DOMAIN      731    921       LAMININ DOMAIN IV 2.
FT   DOMAIN      922    954       LAMININ EGF-LIKE 4 (C-TERMINAL).
FT   DOMAIN      955   1004       LAMININ EGF-LIKE 5.
FT   DOMAIN     1011   1060       LAMININ EGF-LIKE 6.
FT   DOMAIN     1061   1111       LAMININ EGF-LIKE 7.
FT   DOMAIN     1126   1222       IG-LIKE C2-TYPE 3.
FT   DOMAIN     1226   1311       IG-LIKE C2-TYPE 4.
FT   DOMAIN     1319   1401       IG-LIKE C2-TYPE 5.
FT   DOMAIN     1410   1499       IG-LIKE C2-TYPE 6.
FT   DOMAIN     1503   1585       IG-LIKE C2-TYPE 7.
FT   DOMAIN     1588   1680       IG-LIKE C2-TYPE 8.
FT   DOMAIN     1690   1785       IG-LIKE C2-TYPE 9.
FT   DOMAIN     1793   1878       IG-LIKE C2-TYPE 10.
FT   DOMAIN     1886   1970       IG-LIKE C2-TYPE 11.
FT   DOMAIN     1973   2069       IG-LIKE C2-TYPE 12.
FT   DOMAIN     2073   2163       IG-LIKE C2-TYPE 13.
FT   DOMAIN     2173   2260       IG-LIKE C2-TYPE 14.
FT   DOMAIN     2263   2343       IG-LIKE C2-TYPE 15.
FT   DOMAIN     2349   2435       IG-LIKE C2-TYPE 16.
FT   DOMAIN     2446   2530       IG-LIKE C2-TYPE 17.
FT   DOMAIN     2532   2713       LAMININ G-LIKE 1.
FT   DOMAIN     2793   2960       LAMININ G-LIKE 2.
FT   DOMAIN     2961   3093       GLU-RICH.
FT   DOMAIN     2972   3066       THR-RICH.
FT   DOMAIN     3180   3359       LAMININ G-LIKE 3.
FT   DISULFID     66    114       BY SIMILARITY.
FT   DISULFID    149    161       BY SIMILARITY.
FT   DISULFID    156    174       BY SIMILARITY.
FT   DISULFID    168    183       BY SIMILARITY.
FT   DISULFID    190    202       BY SIMILARITY.
FT   DISULFID    197    215       BY SIMILARITY.
FT   DISULFID    209    224       BY SIMILARITY.
FT   DISULFID    233    246       BY SIMILARITY.
FT   DISULFID    240    259       BY SIMILARITY.
FT   DISULFID    253    268       BY SIMILARITY.
FT   DISULFID    955    964       BY SIMILARITY.
FT   DISULFID    957    971       BY SIMILARITY.
FT   DISULFID    974    983       BY SIMILARITY.
FT   DISULFID    986   1002       BY SIMILARITY.
FT   DISULFID   1011   1021       BY SIMILARITY.
FT   DISULFID   1013   1027       BY SIMILARITY.
FT   DISULFID   1030   1039       BY SIMILARITY.
FT   DISULFID   1042   1058       BY SIMILARITY.
FT   DISULFID   1061   1069       BY SIMILARITY.
FT   DISULFID   1063   1079       BY SIMILARITY.
FT   DISULFID   1082   1091       BY SIMILARITY.
FT   DISULFID   1094   1109       BY SIMILARITY.
FT   DISULFID   1152   1200       BY SIMILARITY.
FT   DISULFID   1338   1384       BY SIMILARITY.
FT   DISULFID   1435   1481       BY SIMILARITY.
FT   DISULFID   1527   1573       BY SIMILARITY.
FT   DISULFID   1618   1663       BY SIMILARITY.
FT   DISULFID   1719   1767       BY SIMILARITY.
FT   DISULFID   1814   1861       BY SIMILARITY.
FT   DISULFID   1907   1954       BY SIMILARITY.
```

```
FT   DISULFID    1998   2053       BY SIMILARITY.
FT   DISULFID    2099   2147       BY SIMILARITY.
FT   DISULFID    2195   2242       BY SIMILARITY.
FT   DISULFID    2284   2329       BY SIMILARITY.
FT   DISULFID    2374   2420       BY SIMILARITY.
FT   DISULFID    2467   2514       BY SIMILARITY.
FT   CARBOHYD    1422   1422       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    2476   2476       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    2950   2950       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    3143   3143       N-LINKED (GLCNAC. . .) (POTENTIAL).

  Query Match             30.4%;  Score 338.5;  DB 1;  Length 3375;
  Best Local Similarity   35.6%;  Pred. No. 2.7e-19;
  Matches   68;  Conservative   38;  Mismatches   76;  Indels    9;  Gaps    2;

Qy          18 PGQYGAYFHDDGFLAFPGHVFSRSLPEVPETIELEVRTSTASGLLLWQGVEVGEAGQGKD 77
               | :: | |  | |:      |     |  | :  : :|   :|:||||| :     | :|
Db        3178 PIEHAARFDGDAFIELSSDEFPHLTSEKDEIVAFKFKTEQQNGVLLWQG-QRPTVQQMED 3236

Qy          78 FISLGLQDGHLVFRYQLGSGEARLVSEDPINDGEWHRVTALREGRRGSIQVDGEELVSGR 137
               :||:|: :||| | |:|| | | |:||: ::||: | |   |:|| | :::|      ||
Db        3237 YISVGIVNGHLHFSYELGGGAAHLISEERVDDGKEHSVRFERKGREGQMRIDNYREVDGR 3296

Qy         138 SPGPNVAVNAKGSVYIGGAPDVATLTGGRFSSGITGCVKNLVLHSARPGAPPPQPLDLQH 197
               | |    :|  |::::|| ||::  ||| ||:   ||: :: |:  :        |||
Db        3297 STGILAMLNVDGNIFVGGVPDISKATGGLFSNNFVGCIADVELNGVK--------LDLMA 3348

Qy         198 RAQAGANTRPC 208
                |  | | :||
Db        3349 TAIDGKNVKPC 3359


RESULT 4
AGRI_RAT
ID   AGRI_RAT       STANDARD;      PRT;  1959 AA.
AC   P25304; Q63034;
DT   01-MAY-1992 (Rel. 22, Created)
DT   01-JUL-1993 (Rel. 26, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Agrin precursor.
GN   AGRN.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE OF 1-1779 AND 1799-1959 FROM N.A.
RC   TISSUE=Embryonic spinal cord;
RX   MEDLINE=91222570; PubMed=1851019;
RA   Rupp F., Payan D.G., Magill-Solc C., Cowan D.M., Scheller R.H.;
RT   "Structure and expression of a rat agrin.";
RL   Neuron 6:811-823(1991).
RN   [2]
RP   SEQUENCE OF 1777-1801 FROM N.A.
RX   MEDLINE=92407628; PubMed=1326608;
RA   Rupp F., Oezcelik T., Linial M., Peterson K., Francke U., Scheller R.;
RT   "Structure and chromosomal localization of the mammalian agrin gene.";
RL   J. Neurosci. 12:3535-3544(1992).
CC   -!- FUNCTION: Component of the basal lamina that causes the
CC       aggregation of acetylcholine receptors and acetylcholine-esterase
CC       on the surface of muscle fibers of the neuromuscular junction.
CC   -!- SUBUNIT: Binds to laminin.
CC   -!- SUBCELLULAR LOCATION: Synaptic basal lamina at the neuromuscular
CC       junction.
CC   -!- ALTERNATIVE PRODUCTS:
CC       Event=Alternative splicing; Named isoforms=5;
CC         Comment=Additional isoforms seem to exist. Isoforms differ in
CC         their acetylcholine receptor clustering activity;
CC       Name=1;
CC         IsoId=P25304-1; Sequence=Displayed;
CC       Name=2;
CC         IsoId=P25304-2; Sequence=VSP_001365;
CC       Name=3;
```

```
CC         IsoId=P25304-3; Sequence=VSP_001366;
CC       Name=4;
CC         IsoId=P25304-4; Sequence=VSP_001367;
CC       Name=5;
CC         IsoId=P25304-5; Sequence=VSP_001368;
CC   -!- TISSUE SPECIFICITY: Embryonic nervous system and muscle.
CC   -!- DEVELOPMENTAL STAGE: More abundant early in development.
CC   -!- PTM: Contains heparan sulfate chains as well as N-linked and O-
CC       linked oligosaccharides (By similarity).
CC   -!- SIMILARITY: Contains 9 Kazal-like domains.
CC   -!- SIMILARITY: Contains 2 laminin EGF-like domains.
CC   -!- SIMILARITY: Contains 4 EGF-like domains.
CC   -!- SIMILARITY: Contains 1 SEA domain.
CC   -!- SIMILARITY: Contains 3 laminin G-like domains.
CC   -!- CAUTION: It is uncertain whether Met-1, Met-18 or Met-24 is the
CC       initiator.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; M64780; AAA40703.1; -.
DR   EMBL; M64780; AAA40702.1; ALT_INIT.
DR   EMBL; S44194; AAB23326.1; -.
DR   PIR; JH0399; AGRT.
DR   HSSP; P00740; 1EDM.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR006209; EGF_like.
DR   InterPro; IPR003645; FolN.
DR   InterPro; IPR002350; kazal.
DR   InterPro; IPR002049; Laminin_EGF.
DR   InterPro; IPR001791; Laminin_G.
DR   InterPro; IPR000082; SEA_domain.
DR   Pfam; PF00008; EGF; 4.
DR   Pfam; PF00050; kazal; 9.
DR   Pfam; PF00053; laminin_EGF; 2.
DR   Pfam; PF00054; laminin_G; 3.
DR   Pfam; PF01390; SEA; 1.
DR   PRINTS; PR00011; EGFLAMININ.
DR   SMART; SM00180; EGF_Lam; 2.
DR   SMART; SM00274; FOLN; 8.
DR   SMART; SM00280; KAZAL; 9.
DR   SMART; SM00282; LamG; 3.
DR   SMART; SM00200; SEA; 1.
DR   PROSITE; PS00022; EGF_1; 6.
DR   PROSITE; PS01186; EGF_2; 1.
DR   PROSITE; PS50026; EGF_3; 4.
DR   PROSITE; PS50025; LAM_G_DOMAIN; 3.
DR   PROSITE; PS01248; LAMININ_TYPE_EGF; 1.
DR   PROSITE; PS50024; SEA; 1.
KW   Glycoprotein; EGF-like domain; Repeat; Alternative splicing; Signal;
KW   Laminin EGF-like domain; Proteoglycan; Heparan sulfate.
FT   SIGNAL          1     29       POTENTIAL.
FT   CHAIN          30   1959       AGRIN.
FT   DOMAIN         65    137       KAZAL-LIKE 1.
FT   DOMAIN        141    212       KAZAL-LIKE 2.
FT   DOMAIN        213    284       KAZAL-LIKE 3.
FT   DOMAIN        287    356       KAZAL-LIKE 4.
FT   DOMAIN        361    429       KAZAL-LIKE 5.
FT   DOMAIN        430    494       KAZAL-LIKE 6.
FT   DOMAIN        495    559       KAZAL-LIKE 7.
FT   DOMAIN        563    645       KAZAL-LIKE 8.
FT   DOMAIN        688    741       LAMININ EGF-LIKE 1.
FT   DOMAIN        742    788       LAMININ EGF-LIKE 2.
FT   DOMAIN        794    864       KAZAL-LIKE 9.
FT   DOMAIN       1023   1145       SEA.
FT   DOMAIN       1220   1258       EGF-LIKE 1.
FT   DOMAIN       1263   1439       LAMININ G-LIKE 1.
FT   DOMAIN       1440   1477       EGF-LIKE 2.
```

```
FT   DOMAIN     1479   1516       EGF-LIKE 3.
FT   DOMAIN     1526   1708       LAMININ G-LIKE 2.
FT   DOMAIN     1709   1748       EGF-LIKE 4.
FT   DOMAIN     1784   1956       LAMININ G-LIKE 3.
FT   DOMAIN      869    992       SER/THR-RICH.
FT   DOMAIN     1147   1215       SER/THR-RICH.
FT   DISULFID     97    116       POTENTIAL.
FT   DISULFID    105    137       POTENTIAL.
FT   DISULFID    171    191       POTENTIAL.
FT   DISULFID    180    212       POTENTIAL.
FT   DISULFID    244    263       POTENTIAL.
FT   DISULFID    252    284       POTENTIAL.
FT   DISULFID    316    335       POTENTIAL.
FT   DISULFID    324    356       POTENTIAL.
FT   DISULFID    389    408       POTENTIAL.
FT   DISULFID    397    429       POTENTIAL.
FT   DISULFID    454    473       POTENTIAL.
FT   DISULFID    462    494       POTENTIAL.
FT   DISULFID    518    538       POTENTIAL.
FT   DISULFID    527    559       POTENTIAL.
FT   DISULFID    604    624       POTENTIAL.
FT   DISULFID    613    645       POTENTIAL.
FT   DISULFID    688    700       BY SIMILARITY.
FT   DISULFID    690    707       BY SIMILARITY.
FT   DISULFID    709    718       BY SIMILARITY.
FT   DISULFID    721    739       BY SIMILARITY.
FT   DISULFID    742    754       BY SIMILARITY.
FT   DISULFID    744    761       BY SIMILARITY.
FT   DISULFID    763    772       BY SIMILARITY.
FT   DISULFID    775    786       BY SIMILARITY.
FT   DISULFID    823    843       POTENTIAL.
FT   DISULFID    832    864       POTENTIAL.
FT   DISULFID   1224   1235       BY SIMILARITY.
FT   DISULFID   1229   1246       BY SIMILARITY.
FT   DISULFID   1248   1257       BY SIMILARITY.
FT   DISULFID   1444   1455       POTENTIAL.
FT   DISULFID   1449   1465       POTENTIAL.
FT   DISULFID   1467   1476       POTENTIAL.
FT   DISULFID   1483   1494       BY SIMILARITY.
FT   DISULFID   1488   1504       BY SIMILARITY.
FT   DISULFID   1506   1515       BY SIMILARITY.
FT   DISULFID   1713   1727       BY SIMILARITY.
FT   DISULFID   1721   1736       BY SIMILARITY.
FT   DISULFID   1738   1747       BY SIMILARITY.
FT   CARBOHYD    145    145       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    672    672       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    827    827       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    957    957       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   VARSPLIC   1144   1152       Missing (in isoform 2).
FT                                /FTId=VSP_001365.
FT   VARSPLIC   1780   1798       Missing (in isoform 3).
FT                                /FTId=VSP_001366.
FT   VARSPLIC   1788   1798       Missing (in isoform 4).
FT                                /FTId=VSP_001367.
FT   VARSPLIC   1780   1787       Missing (in isoform 5).
FT                                /FTId=VSP_001368.
FT   VARIANT     314    314       V -> VTCD (IN A VARIANT).
SQ   SEQUENCE   1959 AA;  208645 MW;  7FEFDFDAFF89CC31 CRC64;

  Query Match             21.7%;  Score 242.5;  DB 1;  Length 1959;
  Best Local Similarity   32.6%;  Pred. No. 9.3e-12;
  Matches   61;  Conservative   26;  Mismatches   87;  Indels   13;  Gaps    3;

Qy         25 FHDDGFLAFPGHVFSRSLPEVPETIELEVRTSTASGLLLWQGVEVGEAGQGKDFISLGLQ 84
              |    |||||      :|      : || |     ||||: |       :||||::| |
Db       1268 FKGHSFLAFPTLRAYHTL-----RLALEFRALETEGLLLYNG-----NARGKDFLALALL 1317

Qy         85 DGHLVFRYQLGSGEARLVSEDPINDGEWHRVTALREGRRGSIQVDGEELVSGRSPGPNVA 144
              || : ||:  ||| | | |  |:  | |||:   |  |:|:: ||||  | | ||
Db       1318 DGRVQFRFDTGSGPAVLTSLVPVEPGRWHRLELSRHWRQGTLSVDGETPVVGESPSGTDG 1377

Qy        145 VNAKGSVYIGGAPD---VATLTGGRFSSGITGCVKNLVLHSARPGAPPPQPLDLQHRAQA 201
              :|   ::|:|| |:      |       |: ||:: | ::: :      |   :|
```

```
Db       1378 LNLDTNLYVGGIPEEQVAMVLDRTSVGVGLKGCIRMLDINNQQLELSDWQRAAVQSSGVG 1437

Qy        202 GANTRPC 208
                   ||
Db       1438 ECGDHPC 1444

RESULT 5
AGRI_DISOM
ID   AGRI_DISOM     STANDARD;      PRT;  1328 AA.
AC   Q90404;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Agrin (Fragment).
GN   AGRN.
OS   Discopyge ommata (Electric ray).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Chondrichthyes;
OC   Elasmobranchii; Squalea; Hypnosqualea; Pristiorajea; Batoidea;
OC   Torpediniformes; Narcinoidei; Narcinidae; Discopyge.
OX   NCBI_TaxID=7785;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Smith M.A., Magill-Solc C., Rupp F., Yao Y.-M.M., Schilling J.W.,
RA   Snow P., McMahan U.J.;
RT   "Isolation and characterization of a cDNA that encodes an agrin
RT   homolog in the marine ray.";
RL   Mol. Cell. Neurosci. 3:406-417(1992).
CC   -!- FUNCTION: Component of the basal lamina that causes the
CC       aggregation of acetylcholine receptors and acetylcholine-esterase
CC       on the surface of muscle fibers of the neuromuscular junction (By
CC       similarity).
CC   -!- SUBCELLULAR LOCATION: Synaptic basal lamina at the neuromuscular
CC       junction (By similarity).
CC   -!- SIMILARITY: Contains at least 2 laminin EGF-like domains.
CC   -!- SIMILARITY: Contains at least 3 EGF-like domains.
CC   -!- SIMILARITY: Contains 1 SEA domain.
CC   -!- SIMILARITY: Contains 3 laminin G-like domains.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; L01423; AAA49224.1; -.
DR   PIR; T43060; T43060.
DR   HSSP; P00740; 1EDM.
DR   InterPro; IPR000152; Asx_hydroxyl_S.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR000742; EGF_2.
DR   InterPro; IPR006209; EGF_like.
DR   InterPro; IPR003645; FolN.
DR   InterPro; IPR002350; kazal.
DR   InterPro; IPR002049; Laminin_EGF.
DR   InterPro; IPR001791; Laminin_G.
DR   InterPro; IPR000082; SEA_domain.
DR   Pfam; PF00008; EGF; 4.
DR   Pfam; PF00050; kazal; 2.
DR   Pfam; PF00053; laminin_EGF; 2.
DR   Pfam; PF00054; laminin_G; 3.
DR   Pfam; PF01390; SEA; 1.
DR   PRINTS; PR00011; EGFLAMININ.
DR   SMART; SM00180; EGF_Lam; 2.
DR   SMART; SM00274; FOLN; 2.
DR   SMART; SM00280; KAZAL; 2.
DR   SMART; SM00282; LamG; 3.
DR   SMART; SM00200; SEA; 1.
DR   PROSITE; PS00010; ASX_HYDROXYL; 1.
DR   PROSITE; PS00022; EGF_1; 5.
DR   PROSITE; PS01186; EGF_2; 1.
```

```
DR   PROSITE; PS50026; EGF_3; 4.
DR   PROSITE; PS50025; LAM_G_DOMAIN; 3.
DR   PROSITE; PS01248; LAMININ_TYPE_EGF; 1.
DR   PROSITE; PS50024; SEA; 1.
KW   Glycoprotein; EGF-like domain; Repeat; Laminin EGF-like domain.
FT   NON_TER       1      1
FT   DOMAIN       79    132       LAMININ EGF-LIKE 1.
FT   DOMAIN      133    179       LAMININ EGF-LIKE 2.
FT   DOMAIN      411    533       SEA.
FT   DOMAIN      608    644       EGF-LIKE 1.
FT   DOMAIN      649    825       LAMININ G-LIKE 1.
FT   DOMAIN      865    902       EGF-LIKE 2.
FT   DOMAIN      914   1096       LAMININ G-LIKE 2.
FT   DOMAIN     1097   1135       EGF-LIKE 3.
FT   DOMAIN     1146   1324       LAMININ G-LIKE 3.
FT   DISULFID     79     91       BY SIMILARITY.
FT   DISULFID     81     98       BY SIMILARITY.
FT   DISULFID    100    109       BY SIMILARITY.
FT   DISULFID    112    130       BY SIMILARITY.
FT   DISULFID    133    145       BY SIMILARITY.
FT   DISULFID    135    152       BY SIMILARITY.
FT   DISULFID    154    163       BY SIMILARITY.
FT   DISULFID    166    177       BY SIMILARITY.
FT   DISULFID    612    623       BY SIMILARITY.
FT   DISULFID    617    632       BY SIMILARITY.
FT   DISULFID    634    643       BY SIMILARITY.
FT   DISULFID    869    880       BY SIMILARITY.
FT   DISULFID    874    890       BY SIMILARITY.
FT   DISULFID    892    901       BY SIMILARITY.
FT   DISULFID   1101   1114       BY SIMILARITY.
FT   DISULFID   1108   1123       BY SIMILARITY.
FT   DISULFID   1125   1134       BY SIMILARITY.
FT   CARBOHYD     63     63       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    168    168       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    279    279       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    301    301       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    369    369       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   1328 AA;  144018 MW;  79D81C1AF2A71C18 CRC64;

  Query Match             21.1%;  Score 235;  DB 1;  Length 1328;
  Best Local Similarity   36.9%;  Pred. No. 2.4e-11;
  Matches   62;  Conservative   32;  Mismatches   56;  Indels   18;  Gaps    8;

Qy          25 FHDDGFLAFPG-HVFSRSLPEVPETIELEV--RTSTASGLLLWQGVEVGEAGQGKDFISL 81
               |:   :|   | | |   : :: : : :||       :|:: : | :    |:| ||:||
Db         919 FNGLSYLEMNGIHTF---VSDLLQKLSMEVIFLAKDPNGMIFYNGQKTD--GRG-DFVSL 972

Qy          82 GLQDGHLVFRYQLGSGEARLVSEDPINDGEWHRVTALREGRRGSIQVDGEELVSGRSP-- 139
                |:||:| |:| || | | | |: ||    |: ||  | ||:| :::: :||||| ||
Db         973 NLRDGYLEFKYDLGKGAAVLRSKAPIPLNVWNVVTVERNGRKGLMKINKDELVSGESPKS 1032

Qy         140 --GPNVAVNAKGSVYIGGAPD---VATLTGGRFSSGITGCVKNLVLHS 182
                  |: |:| | : |:|||||    |   |    || || :: | | |
Db        1033 RKAPHTALNLKEAFYVGGAPDFNKFARAAG--IISGFTGAIQKLSLKS 1078

RESULT 6
LMA2_HUMAN
ID   LMA2_HUMAN     STANDARD;      PRT;  3110 AA.
AC   P24043; Q14736; Q93022;
DT   01-MAR-1992 (Rel. 21, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   15-MAR-2004 (Rel. 43, Last annotation update)
DE   Laminin alpha-2 chain precursor (Laminin M chain) (Merosin heavy
DE   chain).
GN   LAMA2 OR LAMM.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Placenta;
```

```
RX   MEDLINE=94124633; PubMed=8294519;
RA   Vuolteenaho R., Nissinen M., Sainio K., Byers M., Eddy R.,
RA   Hirvonen H., Shows T.B., Sariola H., Engvall E., Tryggvason K.;
RT   "Human laminin M chain (merosin): complete primary structure,
RT   chromosomal assignment, and expression of the M and A chain in human
RT   fetal tissues.";
RL   J. Cell Biol. 124:381-394(1994).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=97066955; PubMed=8910357;
RA   Zhang X., Vuolteenaho R., Tryggvason K.;
RT   "Structure of the human laminin alpha2-chain gene (LAMA2), which is
RT   affected in congenital muscular dystrophy.";
RL   J. Biol. Chem. 271:27664-27669(1996).
RN   [3]
RP   SEQUENCE OF 1981-3110 FROM N.A., AND PARTIAL SEQUENCE.
RC   TISSUE=Placenta;
RX   MEDLINE=90238994; PubMed=2185464;
RA   Ehrig K., Leivo I., Argraves W.S., Ruoslahti E., Engvall E.;
RT   "Merosin, a tissue-specific basement membrane protein, is a
RT   laminin-like protein.";
RL   Proc. Natl. Acad. Sci. U.S.A. 87:3264-3268(1990).
RN   [4]
RP   VARIANTS GLN-545; HIS-619; LEU-919; HIS-2586 AND LYS-2614.
RA   Panicker S.G., Mendell J.T., Chen L., Feng B., Sahenk Z.,
RA   Marzluf G.A., Amato A.A., Mendell J.R.;
RT   "Novel single base polymorphisms and rare sequence variants in
RT   the laminin 2-chain coding region detected by RNA/SSCP analysis.";
RL   Hum. Mutat. 13:174-174(1999).
RN   [5]
RP   ERRATUM.
RA   Panicker S.G., Mendell J.T., Chen L., Feng B., Sahenk Z.,
RA   Marzluf G.A., Amato A.A., Mendell J.R.;
RL   Hum. Mutat. 13:340-340(1999).
RN   [6]
RP   VARIANT MDC1A PRO-2564.
RX   MEDLINE=21476011; PubMed=11591858;
RA   He Y., Jones K.J., Vignier N., Morgan G., Chevallay M., Barois A.,
RA   Estournet-Mathiaud B., Hori H., Mizuta T., Tome F.M.S., North K.N.,
RA   Guicheney P.;
RT   "Congenital muscular dystrophy with primary partial laminin alpha-2
RT   chain deficiency: molecular study.";
RL   Neurology 57:1319-1322(2001).
RN   [7]
RP   VARIANTS MDC1A TYR-527 AND ARG-862.
RX   MEDLINE=22439669; PubMed=12552556;
RA   Tezak Z., Prandini P., Boscaro M., Marin A., Devaney J., Marino M.,
RA   Fanin M., Trevisan C.P., Park J., Tyson W., Finkel R., Garcia C.,
RA   Angelini C., Hoffman E.P., Pegoraro E.;
RT   "Clinical and molecular study in congenital muscular dystrophy with
RT   partial laminin alpha-2 (LAMA2) deficiency.";
RL   Hum. Mutat. 21:103-111(2003).
CC   -!- FUNCTION: Binding to cells via a high affinity receptor, laminin
CC       is thought to mediate the attachment, migration and organization
CC       of cells into tissues during embryonic development by interacting
CC       with other extracellular matrix components.
CC   -!- SUBUNIT: Laminin is a complex glycoprotein, consisting of three
CC       different polypeptide chains (alpha, beta, gamma), which are bound
CC       to each other by disulfide bonds into a cross-shaped molecule
CC       comprising one long and three short arms with globules at each
CC       end. The alpha-2 chain is a subunit of laminin-2 (merosin) and
CC       laminin-4 (S-merosin).
CC   -!- SUBCELLULAR LOCATION: Extracellular; found in the basement
CC       membranes (major component).
CC   -!- TISSUE SPECIFICITY: Placenta, striated muscle, peripheral nerve,
CC       cardiac muscle, pancreas, lung, spleen, kidney, adrenal gland,
CC       skin, testis, meninges, choroid plexus, and some other regions of
CC       the brain; not in liver, thymus and bone.
CC   -!- DOMAIN: The alpha-helical domains I and II are thought to interact
CC       with other laminin chains to form a coiled coil structure.
CC   -!- DOMAIN: Domains VI, IV and G are globular.
CC   -!- DISEASE: Defects in LAMA2 are the cause of merosin-deficient
CC       congenital muscular dystrophy type 1A (MDC1A) [MIM:607855]. MDC1A
```

```
CC        is characterized by difficulty walking, hypotonia, proximal
CC        weakness, hyporeflexia, and white matter hypodensity on MRI.
CC    -!- SIMILARITY: Contains 1 laminin N-terminal domain.
CC    -!- SIMILARITY: Contains 17 laminin EGF-like domains.
CC    -!- SIMILARITY: Contains 2 laminin IV domains.
CC    -!- SIMILARITY: Contains 5 laminin G-like domains.
CC    --------------------------------------------------------------------------
CC    This SWISS-PROT entry is copyright. It is produced through a collaboration
CC    between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC    the European Bioinformatics Institute.  There are no  restrictions on  its
CC    use  by  non-profit  institutions as long  as its content  is  in  no  way
CC    modified and this statement is not removed.  Usage  by  and for commercial
CC    entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC    or send an email to license@isb-sib.ch).
CC    --------------------------------------------------------------------------
DR    EMBL; Z26653; CAA81394.1; -.
DR    EMBL; U66796; AAB18388.1; -.
DR    EMBL; U66733; AAB18388.1; JOINED.
DR    EMBL; U66734; AAB18388.1; JOINED.
DR    EMBL; U66735; AAB18388.1; JOINED.
DR    EMBL; U66736; AAB18388.1; JOINED.
DR    EMBL; U66737; AAB18388.1; JOINED.
DR    EMBL; U66738; AAB18388.1; JOINED.
DR    EMBL; U66739; AAB18388.1; JOINED.
DR    EMBL; U66740; AAB18388.1; JOINED.
DR    EMBL; U66741; AAB18388.1; JOINED.
DR    EMBL; U66742; AAB18388.1; JOINED.
DR    EMBL; U66743; AAB18388.1; JOINED.
DR    EMBL; U66745; AAB18388.1; JOINED.
DR    EMBL; U66746; AAB18388.1; JOINED.
DR    EMBL; U66747; AAB18388.1; JOINED.
DR    EMBL; U66748; AAB18388.1; JOINED.
DR    EMBL; U66749; AAB18388.1; JOINED.
DR    EMBL; U66750; AAB18388.1; JOINED.
DR    EMBL; U66751; AAB18388.1; JOINED.
DR    EMBL; U66752; AAB18388.1; JOINED.
DR    EMBL; U66753; AAB18388.1; JOINED.
DR    EMBL; U66754; AAB18388.1; JOINED.
DR    EMBL; U66755; AAB18388.1; JOINED.
DR    EMBL; U66756; AAB18388.1; JOINED.
DR    EMBL; U66757; AAB18388.1; JOINED.
DR    EMBL; U66758; AAB18388.1; JOINED.
DR    EMBL; U66759; AAB18388.1; JOINED.
DR    EMBL; U66760; AAB18388.1; JOINED.
DR    EMBL; U66761; AAB18388.1; JOINED.
DR    EMBL; U66762; AAB18388.1; JOINED.
DR    EMBL; U66763; AAB18388.1; JOINED.
DR    EMBL; U66764; AAB18388.1; JOINED.
DR    EMBL; U66765; AAB18388.1; JOINED.
DR    EMBL; U66766; AAB18388.1; JOINED.
DR    EMBL; U66768; AAB18388.1; JOINED.
DR    EMBL; U66769; AAB18388.1; JOINED.
DR    EMBL; U66770; AAB18388.1; JOINED.
DR    EMBL; U66771; AAB18388.1; JOINED.
DR    EMBL; U66772; AAB18388.1; JOINED.
DR    EMBL; U66773; AAB18388.1; JOINED.
DR    EMBL; U66774; AAB18388.1; JOINED.
DR    EMBL; U66775; AAB18388.1; JOINED.
DR    EMBL; U66776; AAB18388.1; JOINED.
DR    EMBL; U66777; AAB18388.1; JOINED.
DR    EMBL; U66778; AAB18388.1; JOINED.
DR    EMBL; U66779; AAB18388.1; JOINED.
DR    EMBL; U66780; AAB18388.1; JOINED.
DR    EMBL; U66781; AAB18388.1; JOINED.
DR    EMBL; U66782; AAB18388.1; JOINED.
DR    EMBL; U66783; AAB18388.1; JOINED.
DR    EMBL; U66784; AAB18388.1; JOINED.
DR    EMBL; U66785; AAB18388.1; JOINED.
DR    EMBL; U66786; AAB18388.1; JOINED.
DR    EMBL; U66787; AAB18388.1; JOINED.
DR    EMBL; U66788; AAB18388.1; JOINED.
DR    EMBL; U66789; AAB18388.1; JOINED.
DR    EMBL; U66790; AAB18388.1; JOINED.
```

```
DR    EMBL; U66791; AAB18388.1; JOINED.
DR    EMBL; U66792; AAB18388.1; JOINED.
DR    EMBL; U66793; AAB18388.1; JOINED.
DR    EMBL; U66794; AAB18388.1; JOINED.
DR    EMBL; U66795; AAB18388.1; JOINED.
DR    EMBL; M59832; AAA63215.1; -.
DR    PIR; PX0082; MMHUMH.
DR    HSSP; Q60675; 1QU0.
DR    Genew; HGNC:6482; LAMA2.
DR    MIM; 156225; -.
DR    MIM; 607855; -.
DR    GO; GO:0005604; C:basement membrane; TAS.
DR    GO; GO:0005198; F:structural molecule activity; TAS.
DR    GO; GO:0007517; P:muscle development; TAS.
DR    InterPro; IPR008985; ConA_like_lec_gl.
DR    InterPro; IPR006209; EGF_like.
DR    InterPro; IPR008979; Gal_bind_like.
DR    InterPro; IPR000034; Laminin_B.
DR    InterPro; IPR002049; Laminin_EGF.
DR    InterPro; IPR001791; Laminin_G.
DR    InterPro; IPR008211; LamNT.
DR    Pfam; PF00052; laminin_B; 2.
DR    Pfam; PF00053; laminin_EGF; 14.
DR    Pfam; PF00054; laminin_G; 5.
DR    Pfam; PF00055; laminin_Nterm; 1.
DR    PRINTS; PR00011; EGFLAMININ.
DR    ProDom; PD003031; Laminin_B; 1.
DR    SMART; SM00180; EGF_Lam; 15.
DR    SMART; SM00281; LamB; 2.
DR    SMART; SM00282; LamG; 5.
DR    SMART; SM00136; LamNT; 1.
DR    PROSITE; PS00022; EGF_1; 11.
DR    PROSITE; PS01186; EGF_2; 3.
DR    PROSITE; PS01248; LAMININ_TYPE_EGF; 14.
DR    PROSITE; PS50025; LAM_G_DOMAIN; 5.
KW    Glycoprotein; Basement membrane; Extracellular matrix; Coiled coil;
KW    Laminin EGF-like domain; Cell adhesion; Repeat; Signal; Polymorphism;
KW    Disease mutation.
FT    SIGNAL        1     22       POTENTIAL.

  Query Match             20.0%;  Score 223.5;  DB 1;  Length 3110;
  Best Local Similarity   36.4%;  Pred. No. 5.6e-10;
  Matches   59;  Conservative   22;  Mismatches   68;  Indels   13;  Gaps    5;

Qy          48 TIELEVRTSTASGLLLWQGVEVGEAGQGKDFISLGLQDGHLVFRYQLGSGEARLVSEDPI 107
               ||||||||   |||| :       |    || :: |::|   | | ||||:   :    |
Db        2788 TIELEVRTEAESGLLFYMA-----AINHADFATVQLRNGLPYFSYDLGSGDTHTMIPTKI 2842

Qy         108 NDGEWHRVTALREGRRGSIQVDGEELVSGRSPGPNVA--VNAKGSVYIGGAP-DVATLTG 164
               |||:||::  :|  : | : |||    | |:  |  |  ::  | :|:|| | :  |
Db        2843 NDGQWHKIKIMRSKQEGILYVDG---ASNRTISPKKADILDVVGMLYVGGLPINYTTRRI 2899

Qy         165 GRFSSGITGCVKNLVLHSARPGAPPPQPLDLQHRAQAGANTR 206
               |  :  | |||:|  || |   |   ||    |     || :
Db        2900 GPVTYSIDGCVRN--LHMAEAPADLEQPTSSFHVGTCFANAQ 2939


RESULT 7
LMA2_MOUSE
ID   LMA2_MOUSE     STANDARD;      PRT;  3106 AA.
AC   Q60675; Q05003; Q64061;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   15-MAR-2004 (Rel. 43, Last annotation update)
DE   Laminin alpha-2 chain precursor (Laminin M chain) (Merosin heavy
DE   chain).
GN   LAMA2.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
```

```
RC   STRAIN=FVB/N; TISSUE=Embryo, and Heart;
RX   MEDLINE=95316259; PubMed=7795883;
RA   Bernier S.M., Utani A., Sugiyama S., Doi T., Polistina C.,
RA   Yamada Y.;
RT   "Cloning and expression of laminin alpha 2 chain (M-chain) in the
RT   mouse.";
RL   Matrix Biol. 14:447-455(1995).
RN   [2]
RP   SEQUENCE OF 2162-2279 FROM N.A.
RC   STRAIN=C57BL/6; TISSUE=Thymus;
RX   MEDLINE=93346725; PubMed=8345183;
RA   Chang A.C., Wadsworth S., Coligan J.E.;
RT   "Expression of merosin in the thymus and its interaction with
RT   thymocytes.";
RL   J. Immunol. 151:1789-1801(1993).
RN   [3]
RP   SEQUENCE OF 64-281 FROM N.A.
RX   MEDLINE=95179178; PubMed=7874173;
RA   Xu H., Wu X.R., Wewer U.M., Engvall E.;
RT   "Murine muscular dystrophy caused by a mutation in the laminin alpha
RT   2 (Lama2) gene.";
RL   Nat. Genet. 8:297-302(1994).
RN   [4]
RP   SEQUENCE OF 20-25.
RX   MEDLINE=21818471; PubMed=11829758;
RA   Garbe J.H., Gohring W., Mann K., Timpl R., Sasaki T.;
RT   "Complete sequence, recombinant analysis and binding to laminins and
RT   sulphated ligands of the N-terminal domains of laminin alpha3B and
RT   alpha5 chains.";
RL   Biochem. J. 362:213-221(2002).
RN   [5]
RP   BINDING TO FBLN1, FBLN2, AND NID2.
RX   MEDLINE=99146904; PubMed=10022829;
RA   Talts J.F., Andac Z., Goehring W., Brancaccio A., Timpl R.;
RT   "Binding of the G domains of laminin alpha1 and alpha2 chains and
RT   perlecan to heparin, sulfatides, alpha-dystroglycan and several
RT   extracellular matrix proteins.";
RL   EMBO J. 18:863-870(1999).
RN   [6]
RP   X-RAY CRYSTALLOGRAPHY (2.35 ANGSTROMS) OF 2932-3106.
RX   MEDLINE=20085745; PubMed=10619025;
RA   Hohenester E., Tisi D., Talts J.F., Timpl R.;
RT   "The crystal structure of a laminin G-like module reveals the
RT   molecular basis of alpha-dystroglycan binding to laminins, perlecan,
RT   and agrin.";
RL   Mol. Cell 4:783-792(1999).
CC   -!- FUNCTION: Binding to cells via a high affinity receptor, laminin
CC       is thought to mediate the attachment, migration and organization
CC       of cells into tissues during embryonic development by interacting
CC       with other extracellular matrix components.
CC   -!- SUBUNIT: Laminin is a complex glycoprotein, consisting of three
CC       different polypeptide chains (alpha, beta, gamma), which are bound
CC       to each other by disulfide bonds into a cross-shaped molecule
CC       comprising one long and three short arms with globules at each
CC       end. The alpha-2 chain is a subunit of laminin-2 (Merosin) and
CC       laminin-4 (S-merosin). Interacts with FBLN1, FBLN2 and NID2.
CC   -!- SUBCELLULAR LOCATION: Extracellular.
CC   -!- TISSUE SPECIFICITY: Found in the basement membranes (major
CC       component).
CC   -!- DOMAIN: The alpha-helical domains I and II are thought to interact
CC       with other laminin chains to form a coiled coil structure.
CC   -!- DOMAIN: Domains VI, IV and G are globular.
CC   -!- DISEASE: Defects in LAMA2 are a cause of murine muscular dystrophy
CC       (dy2J).
CC   -!- SIMILARITY: Contains 1 laminin N-terminal domain.
CC   -!- SIMILARITY: Contains 17 laminin EGF-like domains.
CC   -!- SIMILARITY: Contains 2 laminin IV domains.
CC   -!- SIMILARITY: Contains 5 laminin G-like domains.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
```

```
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; U12147; AAC52165.1; -.
DR   EMBL; X69869; CAA49502.1; -.
DR   EMBL; S75315; AAB33573.1; -.
DR   PIR; I49077; S53868.
DR   PDB; 1QU0; 03-DEC-99.
DR   PDB; 1DYK; 04-FEB-01.
DR   MGD; MGI:99912; Lama2.
DR   GO; GO:0005604; C:basement membrane; IDA.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR006209; EGF_like.
DR   InterPro; IPR008979; Gal_bind_like.
DR   InterPro; IPR000034; Laminin_B.
DR   InterPro; IPR002049; Laminin_EGF.
DR   InterPro; IPR001791; Laminin_G.
DR   InterPro; IPR008211; LamNT.
DR   Pfam; PF00052; laminin_B; 2.
DR   Pfam; PF00053; laminin_EGF; 14.
DR   Pfam; PF00054; laminin_G; 5.
DR   Pfam; PF00055; laminin_Nterm; 1.
DR   PRINTS; PR00011; EGFLAMININ.
DR   ProDom; PD003031; Laminin_B; 1.
DR   SMART; SM00180; EGF_Lam; 14.
DR   SMART; SM00281; LamB; 2.
DR   SMART; SM00282; LamG; 5.
DR   SMART; SM00136; LamNT; 1.
DR   PROSITE; PS00022; EGF_1; 11.
DR   PROSITE; PS01186; EGF_2; 3.
DR   PROSITE; PS01248; LAMININ_TYPE_EGF; 14.
DR   PROSITE; PS50025; LAM_G_DOMAIN; 5.
KW   Glycoprotein; Basement membrane; Extracellular matrix; Coiled coil;
KW   Laminin EGF-like domain; Cell adhesion; Repeat; Signal; 3D-structure.
FT   SIGNAL        1     19
FT   CHAIN        20   3106       LAMININ ALPHA-2 CHAIN.
FT   DOMAIN       20    282       LAMININ N-TERMINAL (DOMAIN VI).
FT   DOMAIN      283    339       LAMININ EGF-LIKE 1.
FT   DOMAIN      340    409       LAMININ EGF-LIKE 2.
FT   DOMAIN      410    464       LAMININ EGF-LIKE 3.
FT   DOMAIN      465    513       LAMININ EGF-LIKE 4.
FT   DOMAIN      514    523       LAMININ EGF-LIKE 5 (N-TERMINAL).
FT   DOMAIN      524    719       LAMININ DOMAIN IV 1 (DOMAIN IV B).
FT   DOMAIN      720    752       LAMININ EGF-LIKE 5 (C-TERMINAL).
FT   DOMAIN      753    802       LAMININ EGF-LIKE 6.
FT   DOMAIN      803    860       LAMININ EGF-LIKE 7.
FT   DOMAIN      861    913       LAMININ EGF-LIKE 8.
FT   DOMAIN      914    962       LAMININ EGF-LIKE 9.
FT   DOMAIN      963   1009       LAMININ EGF-LIKE 10.
FT   DOMAIN     1010   1055       LAMININ EGF-LIKE 11.
FT   DOMAIN     1056   1101       LAMININ EGF-LIKE 12.
FT   DOMAIN     1102   1161       LAMININ EGF-LIKE 13.
FT   DOMAIN     1162   1171       LAMININ EGF-LIKE 14 (N-TERMINAL).
FT   DOMAIN     1172   1375       LAMININ DOMAIN IV 2 (DOMAIN IV A).
FT   DOMAIN     1376   1415       LAMININ EGF-LIKE 14 (C-TERMINAL).
FT   DOMAIN     1416   1464       LAMININ EGF-LIKE 15.
FT   DOMAIN     1465   1522       LAMININ EGF-LIKE 16.
FT   DOMAIN     1523   1569       LAMININ EGF-LIKE 17.
FT   DOMAIN     1570   2140       DOMAIN II AND I.
FT   DOMAIN     2141   2324       LAMININ G-LIKE 1.
FT   DOMAIN     2336   2517       LAMININ G-LIKE 2.
FT   DOMAIN     2522   2706       LAMININ G-LIKE 3.
FT   DOMAIN     2759   2930       LAMININ G-LIKE 4.
FT   DOMAIN     2929   3106       LAMININ G-LIKE 5.
FT   DOMAIN     1662   1863       COILED COIL (POTENTIAL).
FT   DOMAIN     1923   2146       COILED COIL (POTENTIAL).
FT   DISULFID    283    292       BY SIMILARITY.
FT   DISULFID    285    303       BY SIMILARITY.
FT   DISULFID    305    314       BY SIMILARITY.
FT   DISULFID    317    337       BY SIMILARITY.
FT   DISULFID    340    349       BY SIMILARITY.
FT   DISULFID    342    374       BY SIMILARITY.
```

```
FT   DISULFID    377    386       BY SIMILARITY.
FT   DISULFID    389    407       BY SIMILARITY.
FT   DISULFID    410    422       BY SIMILARITY.
FT   DISULFID    412    438       BY SIMILARITY.
FT   DISULFID    440    449       BY SIMILARITY.
FT   DISULFID    452    462       BY SIMILARITY.
FT   DISULFID    465    478       BY SIMILARITY.
FT   DISULFID    467    482       BY SIMILARITY.
FT   DISULFID    484    493       BY SIMILARITY.
FT   DISULFID    496    511       BY SIMILARITY.
FT   DISULFID    753    762       BY SIMILARITY.
FT   DISULFID    755    769       BY SIMILARITY.
FT   DISULFID    772    781       BY SIMILARITY.
FT   DISULFID    784    800       BY SIMILARITY.
FT   DISULFID    803    818       BY SIMILARITY.
FT   DISULFID    805    828       BY SIMILARITY.
FT   DISULFID    831    840       BY SIMILARITY.
FT   DISULFID    843    858       BY SIMILARITY.
FT   DISULFID    861    875       BY SIMILARITY.
FT   DISULFID    863    882       BY SIMILARITY.
FT   DISULFID    885    894       BY SIMILARITY.
FT   DISULFID    897    911       BY SIMILARITY.
FT   DISULFID    914    926       BY SIMILARITY.
FT   DISULFID    916    933       BY SIMILARITY.
FT   DISULFID    935    944       BY SIMILARITY.
FT   DISULFID    947    960       BY SIMILARITY.
FT   DISULFID    963    975       BY SIMILARITY.
FT   DISULFID    965    981       BY SIMILARITY.
FT   DISULFID    983    992       BY SIMILARITY.
FT   DISULFID    995   1007       BY SIMILARITY.
FT   DISULFID   1010   1019       BY SIMILARITY.
FT   DISULFID   1012   1026       BY SIMILARITY.
FT   DISULFID   1028   1037       BY SIMILARITY.
FT   DISULFID   1040   1053       BY SIMILARITY.
FT   DISULFID   1056   1068       BY SIMILARITY.
FT   DISULFID   1058   1075       BY SIMILARITY.
FT   DISULFID   1077   1086       BY SIMILARITY.
FT   DISULFID   1089   1099       BY SIMILARITY.
FT   DISULFID   1416   1425       BY SIMILARITY.
FT   DISULFID   1418   1432       BY SIMILARITY.

  Query Match             18.6%;  Score 207.5;  DB 1;  Length 3106;
  Best Local Similarity   35.8%;  Pred. No. 1.1e-08;
  Matches   54;  Conservative   25;  Mismatches   53;  Indels   19;  Gaps    6;

Qy          48 TIELEVRTSTASGLLLWQGVEVGEAGQGKDFISLGLQDGHLVFRYQLGSGEARLVSEDPI 107
               ||||||||   |||| : |  :  |    || :: |::|   | | ||||  | :    |
Db        2784 TIELEVRTEAESGLLFYMG-RINHA----DFGTVQLRNGFPFFSYDLGSGSTRTMIPTKI 2838

Qy         108 NDGEWHRVTALREGRRGSIQVDGEELVSGRSPGPNVA--VNAKGSVYIGGAP-DVATLTG 164
               |||:||::  :|  : | : ||     | ::  |  |  ::  | :|:|| | :  |
Db        2839 NDGQWHKIKIVRVKQEGILYVDD---ASSQTISPKKADILDVGGILYVGGLPINYTTRRI 2895

Qy         165 GRFSSGITGCVKNLVLHSARPGAPPPQPLDL 195
               |  :  : |||:|| :  |        |:||
Db        2896 GPVTYSLDGCVRNLHMEQA--------PVDL 2918


RESULT 8
AGRI_CHICK
ID   AGRI_CHICK     STANDARD;      PRT;  1955 AA.
AC   P31696;
DT   01-JUL-1993 (Rel. 26, Created)
DT   01-JUL-1993 (Rel. 26, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Agrin precursor.
GN   AGRN.
OS   Gallus gallus (Chicken).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC   Gallus.
OX   NCBI_TaxID=9031;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Brain;
RX   MEDLINE=92232297; PubMed=1314620;
RA   Tsim K.W.K., Ruegg M.A., Escher G., Kroeger S., McMahan U.J.;
RT   "cDNA that encodes active agrin.";
RL   Neuron 8:677-689(1992).
RN   [2]
RP   ALTERNATIVE SPLICING.
RX   MEDLINE=92232298; PubMed=1314621;
RA   Ruegg M.A., Tsim K.W.K., Horton S.E., Kroeger S., Escher G.,
RA   Gensch E.M., McMahan U.J.;
RT   "The agrin gene codes for a family of basal lamina proteins that
RT   differ in function and distribution.";
RL   Neuron 8:691-699(1992).
CC   -!- FUNCTION: Component of the basal lamina that causes the
CC       aggregation of acetylcholine receptors and acetylcholine-esterase
CC       on the surface of muscle fibers of the neuromuscular junction.
CC   -!- SUBUNIT: Binds to laminin.
CC   -!- SUBCELLULAR LOCATION: Synaptic basal lamina at the neuromuscular
CC       junction.
CC   -!- ALTERNATIVE PRODUCTS:
CC       Event=Alternative splicing; Named isoforms=3;
CC         Comment=Additional isoforms seem to exist. Isoforms differ in
CC         their acetylcholine receptor clustering activity;
CC       Name=1;
CC         IsoId=P31696-1; Sequence=Displayed;
CC       Name=2; Synonyms=Agrin-related protein 1;
CC         IsoId=P31696-2; Sequence=VSP_001370;
CC       Name=3; Synonyms=Agrin-related protein 2;
CC         IsoId=P31696-3; Sequence=VSP_001369, VSP_001370;
CC   -!- PTM: Contains heparan sulfate chains as well as N-linked and O-
CC       linked oligosaccharides (By similarity).
CC   -!- SIMILARITY: Contains 9 Kazal-like domains.
CC   -!- SIMILARITY: Contains 2 laminin EGF-like domains.
CC   -!- SIMILARITY: Contains 4 EGF-like domains.
CC   -!- SIMILARITY: Contains 1 SEA domain.
CC   -!- SIMILARITY: Contains 3 laminin G-like domains.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; M94271; AAA48585.1; -.
DR   EMBL; M97371; AAA48586.1; -.
DR   EMBL; M97372; -; NOT_ANNOTATED_CDS.
DR   PIR; JH0591; AGCH.
DR   HSSP; P00740; 1EDM.
DR   InterPro; IPR004850; Agrin_NtA.
DR   InterPro; IPR000152; Asx_hydroxyl_S.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR000742; EGF_2.
DR   InterPro; IPR006209; EGF_like.
DR   InterPro; IPR003645; FolN.
DR   InterPro; IPR002350; kazal.
DR   InterPro; IPR002049; Laminin_EGF.
DR   InterPro; IPR001791; Laminin_G.
DR   InterPro; IPR000082; SEA_domain.
DR   InterPro; IPR008993; TIMP_like.
DR   Pfam; PF00008; EGF; 4.
DR   Pfam; PF00050; kazal; 9.
DR   Pfam; PF00053; laminin_EGF; 2.
DR   Pfam; PF00054; laminin_G; 3.
DR   Pfam; PF03146; NtA; 1.
DR   Pfam; PF01390; SEA; 1.
DR   PRINTS; PR00011; EGFLAMININ.
DR   SMART; SM00180; EGF_Lam; 2.
DR   SMART; SM00274; FOLN; 8.
DR   SMART; SM00280; KAZAL; 9.
DR   SMART; SM00282; LamG; 3.
```

```
DR   SMART; SM00200; SEA; 1.
DR   PROSITE; PS00010; ASX_HYDROXYL; 1.
DR   PROSITE; PS00022; EGF_1; 6.
DR   PROSITE; PS01186; EGF_2; 1.
DR   PROSITE; PS50026; EGF_3; 4.
DR   PROSITE; PS50025; LAM_G_DOMAIN; 3.
DR   PROSITE; PS01248; LAMININ_TYPE_EGF; 1.
DR   PROSITE; PS50024; SEA; 1.
KW   Glycoprotein; EGF-like domain; Repeat; Alternative splicing; Signal;
KW   Laminin EGF-like domain; Proteoglycan; Heparan sulfate.
FT   SIGNAL        1     38       POTENTIAL.
FT   CHAIN        39   1955       AGRIN.
FT   DOMAIN       54    126       KAZAL-LIKE 1.
FT   DOMAIN      130    201       KAZAL-LIKE 2.
FT   DOMAIN      202    273       KAZAL-LIKE 3.
FT   DOMAIN      276    344       KAZAL-LIKE 4.
FT   DOMAIN      350    418       KAZAL-LIKE 5.
FT   DOMAIN      419    483       KAZAL-LIKE 6.
FT   DOMAIN      484    548       KAZAL-LIKE 7.
FT   DOMAIN      551    633       KAZAL-LIKE 8.
FT   DOMAIN      675    728       LAMININ EGF-LIKE 1.
FT   DOMAIN      729    775       LAMININ EGF-LIKE 2.
FT   DOMAIN      781    851       KAZAL-LIKE 9.
FT   DOMAIN     1026   1148       SEA.
FT   DOMAIN     1229   1265       EGF-LIKE 1.
FT   DOMAIN     1270   1445       LAMININ G-LIKE 1.
FT   DOMAIN     1446   1483       EGF-LIKE 2.
FT   DOMAIN     1485   1522       EGF-LIKE 3.
FT   DOMAIN     1532   1713       LAMININ G-LIKE 2.
FT   DOMAIN     1714   1752       EGF-LIKE 4.
FT   DOMAIN     1776   1952       LAMININ G-LIKE 3.
FT   DOMAIN      856    995       SER/THR-RICH.
FT   DOMAIN     1150   1219       SER/THR-RICH.
FT   DISULFID     86    105       POTENTIAL.
FT   DISULFID     94    126       POTENTIAL.
FT   DISULFID    160    180       POTENTIAL.
FT   DISULFID    169    201       POTENTIAL.
FT   DISULFID    233    252       POTENTIAL.
FT   DISULFID    241    273       POTENTIAL.
FT   DISULFID    304    323       POTENTIAL.
FT   DISULFID    312    344       POTENTIAL.
FT   DISULFID    378    397       POTENTIAL.
FT   DISULFID    386    418       POTENTIAL.
FT   DISULFID    443    462       POTENTIAL.
FT   DISULFID    451    483       POTENTIAL.
FT   DISULFID    507    527       POTENTIAL.
FT   DISULFID    516    548       POTENTIAL.
FT   DISULFID    592    612       POTENTIAL.
FT   DISULFID    601    633       POTENTIAL.
FT   DISULFID    675    687       BY SIMILARITY.
FT   DISULFID    677    694       BY SIMILARITY.
FT   DISULFID    696    705       BY SIMILARITY.
FT   DISULFID    708    726       BY SIMILARITY.
FT   DISULFID    729    741       BY SIMILARITY.
FT   DISULFID    731    748       BY SIMILARITY.
FT   DISULFID    750    759       BY SIMILARITY.
FT   DISULFID    762    773       BY SIMILARITY.
FT   DISULFID    810    830       POTENTIAL.
FT   DISULFID    819    851       POTENTIAL.
FT   DISULFID   1233   1244       BY SIMILARITY.
FT   DISULFID   1238   1253       BY SIMILARITY.
FT   DISULFID   1255   1264       BY SIMILARITY.
FT   DISULFID   1450   1461       BY SIMILARITY.
FT   DISULFID   1455   1471       BY SIMILARITY.
FT   DISULFID   1473   1482       BY SIMILARITY.
FT   DISULFID   1489   1500       BY SIMILARITY.
FT   DISULFID   1494   1510       BY SIMILARITY.
FT   DISULFID   1512   1521       BY SIMILARITY.
FT   DISULFID   1718   1731       BY SIMILARITY.
FT   DISULFID   1725   1740       BY SIMILARITY.
FT   DISULFID   1742   1751       BY SIMILARITY.
FT   CARBOHYD    390    390       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    659    659       N-LINKED (GLCNAC. . .) (POTENTIAL).
```

```
FT   CARBOHYD    764    764       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    814    814       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   VARSPLIC   1648   1651       Missing (in isoform 3).
FT                                /FTId=VSP_001369.
FT   VARSPLIC   1783   1793       Missing (in isoform 2 and isoform 3).
FT                                /FTId=VSP_001370.
FT   CONFLICT   1129   1131       RTI -> SIL (IN REF. 1; AAA48586).
SQ   SEQUENCE   1955 AA;  211411 MW;  B4DEB27C23422581 CRC64;

  Query Match             18.0%;  Score 200.5;  DB 1;  Length 1955;
  Best Local Similarity   32.9%;  Pred. No. 2.4e-08;
  Matches   56;  Conservative   30;  Mismatches   65;  Indels   19;  Gaps    8;

Qy         48 TIELEVRTSTASGLLLWQGVEVGEAGQGKDFISLGLQDGHLVFRYQLGSGEARLVSEDPI 107
              ::|:     : ||:: : | :    |:| ||:|| | ||:| :|| || | | | |::|:
Db       1559 SMEVVFLAKSPSGMIFYNGQKTD--GKG-DFVSLALHDGYLEYRYDLGKGAAVLRSKEPV 1615

Qy        108 NDGEWHRVTALREGRRGSIQVDGEELVSGRSPG----PNVAVNAKGSVYIGGAPDVATLT 163
                  |  |   | ||:| ::::  | | | ||     |:  :| |   |:||||| : |
Db       1616 PLNTWISVLLERSGRKGVMRINNGERVMGESPKSRKVPHAFLNLKEPFYVGGAPDFSKLA 1675

Qy        164 -GGRFSSGITGCVKNLVLHSARPGAPPPQPLDLQH-RAQAGANT---RPC 208
                   |:   | |: : :     | |    |  || |:    :|    ||
Db       1676 RAAAISTSFYGAVQRISI----KGVP---LLKEQHIRSAVEISTFRAHPC 1718


RESULT 9
NX2A_HUMAN
ID   NX2A_HUMAN     STANDARD;      PRT;  1712 AA.
AC   Q9P2S2; Q9Y2D6;
DT   28-FEB-2003 (Rel. 41, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   15-MAR-2004 (Rel. 43, Last annotation update)
DE   Neurexin 2-alpha precursor (Neurexin II-alpha).
GN   NRXN2 OR KIAA0921.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A. (ISOFORM 1).
RC   TISSUE=Fetal brain;
RA   Seki N., Yoshikawa T., Azuma T., Muramatsu M., Saito T.;
RT   "Human neurexin II.";
RL   Submitted (NOV-1999) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A. (ISOFORM 1).
RA   Rowen L., Madan A., Qin S., Baradarani L., Birditt B., Bloom S.,
RA   Burke J., Dors M., Fleetwood P., Kaur A., Madan A., Nesbitt R.,
RA   Pate D., Hood L.;
RT   "Sequencing of human neurexin II gene.";
RL   Submitted (JUN-2001) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A. (ISOFORM 2), AND REVISIONS.
RC   TISSUE=Brain;
RX   MEDLINE=22158633; PubMed=12168954;
RA   Nakajima D., Okazaki N., Yamakawa H., Kikuno R., Ohara O., Nagase T.;
RT   "Construction of expression-ready cDNA clones for KIAA genes: manual
RT   curation of 330 KIAA cDNA clones.";
RL   DNA Res. 9:99-106(2002).
RN   [4]
RP   SEQUENCE OF 272-1712 FROM N.A. (ISOFORM 2).
RC   TISSUE=Brain;
RX   MEDLINE=99246063; PubMed=10231032;
RA   Nagase T., Ishikawa K.-I., Suyama M., Kikuno R., Hirosawa M.,
RA   Miyajima N., Tanaka A., Kotani H., Nomura N., Ohara O.;
RT   "Prediction of the coding sequences of unidentified human genes. XIII.
RT   The complete sequences of 100 new cDNA clones from brain which code
RT   for large proteins in vitro.";
RL   DNA Res. 6:63-70(1999).
CC   -!- FUNCTION: Neuronal cell surface protein that may be involved in
CC       cell recognition and cell adhesion. May mediate intracellular
CC       signaling.
```

```
CC   -!- SUBUNIT: The laminin G-like domain 1 binds to neurexophilin 1.
CC       Specific isoforms bind to alpha-dystroglycan (By similarity).
CC   -!- SUBCELLULAR LOCATION: Type I membrane protein (Potential).
CC   -!- ALTERNATIVE PRODUCTS:
CC       Event=Alternative promoter;
CC         Comment=A number of isoforms, alpha-type (shown here) and
CC         beta-type (AC P58401), are produced by use of alternative
CC         promoters. Beta-type isoforms differ from alpha-type isoforms
CC         in their N-terminus;
CC       Event=Alternative splicing; Named isoforms=2;
CC         Comment=Additional isoforms seem to exist;
CC       Name=1;
CC         IsoId=Q9P2S2-1; Sequence=Displayed;
CC       Name=2; Synonyms=Alpha-2B;
CC         IsoId=Q9P2S2-2; Sequence=VSP_003505, VSP_003506, VSP_003507,
CC                                  VSP_003508;
CC   -!- TISSUE SPECIFICITY: Predominantly expressed in brain.
CC   -!- SIMILARITY: Contains 6 laminin G-like domains.
CC   -!- SIMILARITY: Contains 3 EGF-like domains.
CC   -!- SIMILARITY: Belongs to the neurexin family.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; AB035266; BAA94075.1; -.
DR   EMBL; AC044790; AAK68154.1; -.
DR   EMBL; AB023138; BAA76765.2; ALT_INIT.
DR   HSSP; Q63373; 1C4R.
DR   Genew; HGNC:8009; NRXN2.
DR   MIM; 600566; -.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR006209; EGF_like.
DR   InterPro; IPR006210; IEGF.
DR   InterPro; IPR001791; Laminin_G.
DR   InterPro; IPR003585; Neurexin-like.
DR   Pfam; PF00008; EGF; 3.
DR   Pfam; PF00054; laminin_G; 5.
DR   SMART; SM00294; 4.1m; 1.
DR   SMART; SM00181; EGF; 3.
DR   SMART; SM00282; LamG; 6.
DR   PROSITE; PS00022; EGF_1; FALSE_NEG.
DR   PROSITE; PS01186; EGF_2; FALSE_NEG.
DR   PROSITE; PS50026; EGF_3; 3.
DR   PROSITE; PS50025; LAM_G_DOMAIN; 6.
KW   Signal; Transmembrane; Repeat; Cell adhesion; EGF-like domain;
KW   Glycoprotein; Alternative splicing; Alternative promoter usage.
FT   SIGNAL        1     28       POTENTIAL.
FT   CHAIN        29   1712       NEUREXIN 2-ALPHA.
FT   DOMAIN       29   1636       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   1637   1657       POTENTIAL.
FT   DOMAIN     1658   1712       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN       29    206       LAMININ G-LIKE 1.
FT   DOMAIN      202    242       EGF-LIKE 1.
FT   DOMAIN      289    486       LAMININ G-LIKE 2.
FT   DOMAIN      493    686       LAMININ G-LIKE 3.
FT   DOMAIN      690    727       EGF-LIKE 2.
FT   DOMAIN      732    904       LAMININ G-LIKE 4.
FT   DOMAIN      918   1093       LAMININ G-LIKE 5.
FT   DOMAIN     1096   1133       EGF-LIKE 3.
FT   DOMAIN     1137   1345       LAMININ G-LIKE 6.
FT   DOMAIN     1368   1371       POLY-THR.
FT   DOMAIN     1374   1377       POLY-THR.
FT   DOMAIN     1444   1447       POLY-PRO.
FT   DOMAIN     1644   1647       POLY-ALA.
FT   CARBOHYD     60     60       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    338    338       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    841    841       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1236   1236       N-LINKED (GLCNAC. . .) (POTENTIAL).
```

```
FT   VARSPLIC    260    283       Missing (in isoform 2).
FT                                /FTId=VSP_003505.
FT   VARSPLIC    393    399       Missing (in isoform 2).
FT                                /FTId=VSP_003506.
FT   VARSPLIC    797    806       DCLRVGCAPS -> G (in isoform 2).
FT                                /FTId=VSP_003507.
FT   VARSPLIC   1253   1282       Missing (in isoform 2).
FT                                /FTId=VSP_003508.
SQ   SEQUENCE   1712 AA;  184980 MW;  E4E8EC404DA1D0B0 CRC64;

  Query Match             17.1%;  Score 191;  DB 1;  Length 1712;
  Best Local Similarity   30.9%;  Pred. No. 1.2e-07;
  Matches   55;  Conservative   30;  Mismatches   53;  Indels   40;  Gaps    8;

Qy         28 DGFLAFPGHVFSRSLPEVPETIELEVRTSTASGLLLW-QGVEVG------EAGQGKDFIS 80
              : |:| |     |:      :| |: ||:  :||||: ||   |       : |  |: :
Db        501 EAFVALPRWSAKRT-----GSISLDFRTTEPNGLLLFSQGRRAGGGAGSHSSAQRADYFA 555

Qy         81 LGLQDGHLVFRYQLGSGEARL-VSEDPINDGEWHRVTALREGRRGSIQVDGEELVSGRSP 139
              : | ||||     :|||  :|  |   :|||||  |   |:||:||| |:     |  :|
Db        556 MELLDGHLYLLLDMGSGGIKLRASSRKVNDGEWCHVDFQRDGRKGSISVN-----SRSTP 610

Qy        140 ----GPNVAVNAKGSVYIGGAPDVATLTGGR-------------FSSGITGCVKNLVL 180
                  | :  :: :  :|:|| |:     |||               :|  |||::| :
Db        611 FLATGDSEILDLESELYLGGLPE-----GGRVDLPLPPEVWTAALRAGYVGCVRDLFI 663

RESULT 10
LMA1_HUMAN
ID   LMA1_HUMAN     STANDARD;      PRT;  3075 AA.
AC   P25391;
DT   01-MAY-1992 (Rel. 22, Created)
DT   01-MAY-1992 (Rel. 22, Last sequence update)
DT   15-MAR-2004 (Rel. 43, Last annotation update)
DE   Laminin alpha-1 chain precursor (Laminin A chain).
GN   LAMA1 OR LAMA.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=91333420; PubMed=1714537;
RA   Haaparanta T., Uitto J., Ruoslahti E., Engvall E.;
RT   "Molecular cloning of the cDNA encoding human laminin A chain.";
RL   Matrix 11:151-160(1991).
RN   [2]
RP   SEQUENCE OF 1-2628 FROM N.A.
RX   MEDLINE=91264789; PubMed=2049067;
RA   Nissinen M., Vuolteenaho R., Boot-Handford R., Kallunki P.,
RA   Tryggvason K.;
RT   "Primary structure of the human laminin A chain. Limited expression
RT   in human tissues.";
RL   Biochem. J. 276:369-379(1991).
RN   [3]
RP   SEQUENCE OF 2397-3072 FROM N.A.
RX   MEDLINE=89280632; PubMed=2733383;
RA   Olsen D., Nagayoshi T., Fazio M., Peltonen J., Jaakkola S.,
RA   Sanborn D., Sasaki T., Kuivaniemi H., Chu M.L., Deutzmann R.,
RA   Timpl R., Uitto J.;
RT   "Human laminin: cloning and sequence analysis of cDNAs encoding A, B1
RT   and B2 chains, and expression of the corresponding genes in human
RT   skin and cultured cells.";
RL   Lab. Invest. 60:772-782(1989).
CC   -!- FUNCTION: Binding to cells via a high affinity receptor, laminin
CC       is thought to mediate the attachment, migration and organization
CC       of cells into tissues during embryonic development by interacting
CC       with other extracellular matrix components.
CC   -!- SUBUNIT: Laminin is a complex glycoprotein, consisting of three
CC       different polypeptide chains (alpha, beta, gamma), which are bound
CC       to each other by disulfide bonds into a cross-shaped molecule
CC       comprising one long and three short arms with globules at each
CC       end. The alpha-1 chain is a subunit of laminin-1 (EHS laminin) and
```

```
CC       laminin-3 (S-laminin).
CC   -!- SUBCELLULAR LOCATION: Extracellular.
CC   -!- TISSUE SPECIFICITY: Found in the basement membranes (major
CC       component).
CC   -!- DOMAIN: The alpha-helical domains I and II are thought to interact
CC       with other laminin chains to form a coiled coil structure.
CC   -!- DOMAIN: Domains VI, IV and G are globular.
CC   -!- SIMILARITY: Contains 1 laminin N-terminal domain.
CC   -!- SIMILARITY: Contains 17 laminin EGF-like domains.
CC   -!- SIMILARITY: Contains 2 laminin IV domains.
CC   -!- SIMILARITY: Contains 5 laminin G-like domains.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; X58531; CAA41418.1; -.
DR   PIR; S14458; S14458.
DR   HSSP; Q60675; 1QU0.
DR   Genew; HGNC:6481; LAMA1.
DR   MIM; 150320; -.
DR   GO; GO:0005606; C:laminin-1; NAS.
DR   GO; GO:0005608; C:laminin-3; NAS.
DR   GO; GO:0007155; P:cell adhesion; NAS.
DR   GO; GO:0016477; P:cell migration; NAS.
DR   GO; GO:0007275; P:development; NAS.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR006209; EGF_like.
DR   InterPro; IPR008979; Gal_bind_like.
DR   InterPro; IPR008212; Lam_N2.
DR   InterPro; IPR000034; Laminin_B.
DR   InterPro; IPR002049; Laminin_EGF.
DR   InterPro; IPR001791; Laminin_G.
DR   InterPro; IPR008211; LamNT.
DR   Pfam; PF00052; laminin_B; 2.
DR   Pfam; PF00053; laminin_EGF; 14.
DR   Pfam; PF00054; laminin_G; 5.
DR   Pfam; PF00055; laminin_Nterm; 1.
DR   PRINTS; PR00011; EGFLAMININ.
DR   ProDom; PD002082; Lam_N2; 1.
DR   ProDom; PD003031; Laminin_B; 2.
DR   SMART; SM00180; EGF_Lam; 15.
DR   SMART; SM00281; LamB; 2.
DR   SMART; SM00282; LamG; 5.
DR   SMART; SM00136; LamNT; 1.
DR   PROSITE; PS00022; EGF_1; 11.
DR   PROSITE; PS01186; EGF_2; 2.
DR   PROSITE; PS01248; LAMININ_TYPE_EGF; 15.
DR   PROSITE; PS50025; LAM_G_DOMAIN; 5.
KW   Glycoprotein; Basement membrane; Extracellular matrix; Coiled coil;
KW   Laminin EGF-like domain; Cell adhesion; Repeat; Signal.
FT   SIGNAL        1     17       POTENTIAL.
FT   CHAIN        18   3075       LAMININ ALPHA-1 CHAIN.
FT   DOMAIN       18    269       LAMININ N-TERMINAL (DOMAIN VI).
FT   DOMAIN      270    326       LAMININ EGF-LIKE 1.
FT   DOMAIN      327    396       LAMININ EGF-LIKE 2.
FT   DOMAIN      397    453       LAMININ EGF-LIKE 3.
FT   DOMAIN      454    502       LAMININ EGF-LIKE 4.
FT   DOMAIN      503    512       LAMININ EGF-LIKE 5 (N-TERMINAL).
FT   DOMAIN      517    708       LAMININ DOMAIN IV 1 (DOMAIN IV B).
FT   DOMAIN      709    741       LAMININ EGF-LIKE 5 (C-TERMINAL).
FT   DOMAIN      742    790       LAMININ EGF-LIKE 6.
FT   DOMAIN      791    848       LAMININ EGF-LIKE 7.
FT   DOMAIN      849    901       LAMININ EGF-LIKE 8.
FT   DOMAIN      902    950       LAMININ EGF-LIKE 9.
FT   DOMAIN      951    997       LAMININ EGF-LIKE 10.
FT   DOMAIN      998   1043       LAMININ EGF-LIKE 11.
FT   DOMAIN     1044   1089       LAMININ EGF-LIKE 12.
FT   DOMAIN     1090   1149       LAMININ EGF-LIKE 13.
```

```
FT   DOMAIN     1150   1159       LAMININ EGF-LIKE 14 (N-TERMINAL).
FT   DOMAIN     1160   1361       LAMININ DOMAIN IV 2 (DOMAIN IV A).
FT   DOMAIN     1362   1402       LAMININ EGF-LIKE 14 (C-TERMINAL).
FT   DOMAIN     1403   1451       LAMININ EGF-LIKE 15.
FT   DOMAIN     1452   1508       LAMININ EGF-LIKE 16.
FT   DOMAIN     1509   1555       LAMININ EGF-LIKE 17.
FT   DOMAIN     1556   2116       DOMAIN II AND I.
FT   DOMAIN     2117   2297       LAMININ G-LIKE 1.
FT   DOMAIN     2305   2481       LAMININ G-LIKE 2.
FT   DOMAIN     2486   2673       LAMININ G-LIKE 3.
FT   DOMAIN     2713   2885       LAMININ G-LIKE 4.
FT   DOMAIN     2890   3070       LAMININ G-LIKE 5.
FT   DOMAIN     1706   1796       COILED COIL (POTENTIAL).
FT   DOMAIN     1968   1989       COILED COIL (POTENTIAL).
FT   DOMAIN     2088   2120       COILED COIL (POTENTIAL).
FT   SITE       2534   2536       CELL ATTACHMENT SITE.
FT   DISULFID    270    279       BY SIMILARITY.
FT   DISULFID    272    290       BY SIMILARITY.
FT   DISULFID    292    301       BY SIMILARITY.
FT   DISULFID    297    305       POTENTIAL.
FT   DISULFID    304    324       BY SIMILARITY.
FT   DISULFID    327    336       BY SIMILARITY.
FT   DISULFID    329    361       BY SIMILARITY.
FT   DISULFID    364    373       BY SIMILARITY.
FT   DISULFID    376    394       BY SIMILARITY.
FT   DISULFID    397    409       BY SIMILARITY.
FT   DISULFID    399    427       BY SIMILARITY.
FT   DISULFID    429    438       BY SIMILARITY.
FT   DISULFID    441    451       BY SIMILARITY.
FT   DISULFID    454    467       BY SIMILARITY.
FT   DISULFID    456    471       BY SIMILARITY.
FT   DISULFID    473    482       BY SIMILARITY.
FT   DISULFID    485    500       BY SIMILARITY.
FT   DISULFID    742    751       BY SIMILARITY.
FT   DISULFID    744    757       BY SIMILARITY.
FT   DISULFID    760    769       BY SIMILARITY.
FT   DISULFID    772    788       BY SIMILARITY.
FT   DISULFID    791    806       BY SIMILARITY.
FT   DISULFID    793    816       BY SIMILARITY.
FT   DISULFID    819    828       BY SIMILARITY.
FT   DISULFID    831    846       BY SIMILARITY.
FT   DISULFID    849    863       BY SIMILARITY.
FT   DISULFID    851    870       BY SIMILARITY.
FT   DISULFID    873    882       BY SIMILARITY.
FT   DISULFID    885    899       BY SIMILARITY.
FT   DISULFID    902    914       BY SIMILARITY.
FT   DISULFID    904    921       BY SIMILARITY.
FT   DISULFID    923    932       BY SIMILARITY.
FT   DISULFID    935    948       BY SIMILARITY.
FT   DISULFID    951    963       BY SIMILARITY.
FT   DISULFID    953    969       BY SIMILARITY.
FT   DISULFID    971    980       BY SIMILARITY.
FT   DISULFID    983    995       BY SIMILARITY.
FT   DISULFID    998   1007       BY SIMILARITY.
FT   DISULFID   1000   1014       BY SIMILARITY.
FT   DISULFID   1016   1025       BY SIMILARITY.
FT   DISULFID   1028   1041       BY SIMILARITY.
FT   DISULFID   1044   1056       BY SIMILARITY.
FT   DISULFID   1046   1063       BY SIMILARITY.
FT   DISULFID   1065   1074       BY SIMILARITY.
FT   DISULFID   1077   1087       BY SIMILARITY.
FT   DISULFID   1403   1412       BY SIMILARITY.
FT   DISULFID   1405   1419       BY SIMILARITY.
FT   DISULFID   1422   1431       BY SIMILARITY.
FT   DISULFID   1434   1449       BY SIMILARITY.
FT   DISULFID   1452   1466       BY SIMILARITY.
FT   DISULFID   1454   1476       BY SIMILARITY.
FT   DISULFID   1479   1488       BY SIMILARITY.
FT   DISULFID   1491   1506       BY SIMILARITY.
FT   DISULFID   1509   1521       BY SIMILARITY.
FT   DISULFID   1511   1528       BY SIMILARITY.
FT   DISULFID   1530   1539       BY SIMILARITY.
FT   DISULFID   1542   1553       BY SIMILARITY.
```

```
FT   DISULFID   1556   1556       INTERCHAIN (PROBABLE).
FT   DISULFID   1560   1560       INTERCHAIN (PROBABLE).
FT   CARBOHYD     38     38       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    555    555       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    665    665       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    763    763       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    926    926       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    952    952       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1045   1045       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1407   1407       N-LINKED (GLCNAC. . .) (POTENTIAL).

  Query Match             16.9%;  Score 188.5;  DB 1;  Length 3075;
  Best Local Similarity   30.5%;  Pred. No. 4e-07;
  Matches   54;  Conservative   30;  Mismatches   70;  Indels   23;  Gaps    7;

Qy         11 GSGGNDAPGQYGAYFHDDGFLAFPGH-VFSRSLPEVPETIELEVRTSTASGLLLWQGVEV 69
              ||  |: |      || ||     |: |  :|||     | :   | : :||||:    :
Db       2299 GSSQNEDPS-----FHFDG----SGYSVVEKSLPATVTQIIMLFNTFSPNGLLLY----L 2345

Qy         70 GEAGQGKDFISLGLQDGHLVFRYQLGSGEARLVSEDPINDGEWHRVTALREGRRGSIQV- 128
              |  |  |||:|: |  | :     ||||   |:::   |:| |:::   |  ::| : |
Db       2346 GSYGT-KDFLSIELFRGRVKVMTDLGSGPITLLTDRRYNNGTWYKIAFQRNRKQGVLAVI 2404

Qy        129 -----DGEELVSGRSPGPNVAVNA--KGSVYIGGAPDVATLTGGRFSSGITGCVKNL 178
                     :|   | :|| :  :|   |  :|:|| |    :  |  :    ||:|||
Db       2405 DAYNTSNKETKQGETPGASSDLNRLDKDPIYVGGLPRSRVVRRGVTTKSFVGCIKNL 2461


RESULT 11
NX2A_RAT
ID   NX2A_RAT        STANDARD;      PRT;  1715 AA.
AC   Q63374; Q63375;
DT   28-FEB-2003 (Rel. 41, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   15-MAR-2004 (Rel. 43, Last annotation update)
DE   Neurexin 2-alpha precursor (Neurexin II-alpha).
GN   NRXN2.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A., VARIANT LEU-434, AND ALTERNATIVE SPLICING.
RC   STRAIN=Sprague-Dawley; TISSUE=Brain;
RX   MEDLINE=92320296; PubMed=1621094;
RA   Ushkaryov Y.A., Petrenko A.G., Geppert M., Suedhof T.C.;
RT   "Neurexins: synaptic cell surface proteins related to the alpha-
RT   latrotoxin receptor and laminin.";
RL   Science 257:50-56(1992).
RN   [2]
RP   SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.
RC   TISSUE=Brain;
RX   MEDLINE=95209856; PubMed=7695896;
RA   Ullrich B., Ushkaryov Y.A., Suedhof T.C.;
RT   "Cartography of neurexins: more than 1000 isoforms generated by
RT   alternative splicing and expressed in distinct subsets of neurons.";
RL   Neuron 14:497-507(1995).
RN   [3]
RP   INTERACTION WITH NEUREXOPHILIN 1.
RX   MEDLINE=99074239; PubMed=9856994;
RA   Missler M., Hammer R.E., Suedhof T.C.;
RT   "Neurexophilin binding to alpha-neurexins. A single LNS domain
RT   functions as an independently folding ligand-binding unit.";
RL   J. Biol. Chem. 273:34716-34723(1998).
RN   [4]
RP   INTERACTION WITH ALPHA-DYSTROGLYCAN.
RX   MEDLINE=21363578; PubMed=11470830;
RA   Sugita S., Saito F., Tang J., Satz J., Campbell K., Suedhof T.C.;
RT   "A stoichiometric complex of neurexins and dystroglycan in brain.";
RL   J. Cell Biol. 154:435-445(2001).
CC   -!- FUNCTION: Neuronal cell surface protein that may be involved in
CC       cell recognition and cell adhesion. May mediate intracellular
CC       signaling.
CC   -!- SUBUNIT: The laminin G-like domain 1 binds to neurexophilin 1.
CC       Isoforms alpha 2C bind to alpha-dystroglycan.
CC   -!- SUBCELLULAR LOCATION: Type I membrane protein (Potential).
CC   -!- ALTERNATIVE PRODUCTS:
CC       Event=Alternative promoter;
CC         Comment=A number of isoforms, alpha-type (shown here) and
CC         beta-type, are produced by use of alternative promoters.
CC         Beta-type isoforms differ from alpha-type isoforms in their
CC         N-terminus;
CC       Event=Alternative splicing; Named isoforms=1;
CC         Comment=At least 216 isoforms may be produced by alternative
CC         splicing. There is a combination of five alternatively spliced
CC         extracellular domains at sites 1 to 5, each consisting of
CC         modular sequences (A-C) that seem to be used independently.
CC         Additional isoforms may derive from a minor cytoplasmic splice
CC         site 6. Beta isoforms (AC Q63376) share the combination of
CC         alternatively spliced domains at sites 4, 5 and 6;
CC       Name=Alpha 1A2A3A4A5A;
CC         IsoId=Q63374-1; Sequence=Displayed;
CC   -!- TISSUE SPECIFICITY: Brain (neuronal synapse).
CC   -!- SIMILARITY: Contains 6 laminin G-like domains.
CC   -!- SIMILARITY: Contains 3 EGF-like domains.
CC   -!- SIMILARITY: Belongs to the neurexin family.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; M96376; AAA41706.1; -.
DR   EMBL; M96376; AAA41707.1; -.
DR   PIR; C40228; C40228.
DR   HSSP; Q63373; 1C4R.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR006209; EGF_like.
DR   InterPro; IPR006210; IEGF.
DR   InterPro; IPR001791; Laminin_G.
DR   InterPro; IPR003585; Neurexin-like.
DR   Pfam; PF00008; EGF; 3.
DR   Pfam; PF00054; laminin_G; 5.
DR   SMART; SM00294; 4.1m; 1.
DR   SMART; SM00181; EGF; 3.
DR   SMART; SM00282; LamG; 6.
DR   PROSITE; PS00022; EGF_1; FALSE_NEG.
DR   PROSITE; PS01186; EGF_2; FALSE_NEG.
DR   PROSITE; PS50026; EGF_3; 3.
DR   PROSITE; PS50025; LAM_G_DOMAIN; 6.
KW   Signal; Transmembrane; Repeat; Cell adhesion; EGF-like domain;
KW   Glycoprotein; Alternative splicing; Alternative promoter usage;
KW   Polymorphism.
FT   SIGNAL        1     29       BY SIMILARITY.
FT   CHAIN        30   1715       NEUREXIN 2-ALPHA.
FT   DOMAIN       30   1639       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   1640   1660       POTENTIAL.
FT   DOMAIN     1661   1715       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN       30    206       LAMININ G-LIKE 1.
FT   DOMAIN      202    242       EGF-LIKE 1.
FT   DOMAIN      289    486       LAMININ G-LIKE 2.
FT   DOMAIN      493    686       LAMININ G-LIKE 3.
FT   DOMAIN      690    727       EGF-LIKE 2.
FT   DOMAIN      732    907       LAMININ G-LIKE 4.
FT   DOMAIN      921   1096       LAMININ G-LIKE 5.
FT   DOMAIN     1099   1136       EGF-LIKE 3.
FT   DOMAIN     1140   1348       LAMININ G-LIKE 6.
FT   DOMAIN     1371   1374       POLY-THR.
FT   DOMAIN     1377   1380       POLY-THR.
FT   DOMAIN     1447   1450       POLY-PRO.
FT   DOMAIN     1647   1650       POLY-ALA.
FT   CARBOHYD     60     60       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    295    295       N-LINKED (GLCNAC. . .) (POTENTIAL).
```

```
FT   CARBOHYD    338    338       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    844    844       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1239   1239       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   VARSPLIC    250    259       Missing (in some isoforms Alpha 1B).
FT                                /FTId=VSP_003509.
FT   VARSPLIC    250    283       Missing (in some isoforms Alpha 1C).
FT                                /FTId=VSP_003510.
FT   VARSPLIC    385    399       Missing (in some isoforms Alpha 2C).
FT                                /FTId=VSP_003511.
FT   VARSPLIC    393    399       Missing (in some isoforms Alpha 2B).
FT                                /FTId=VSP_003512.
FT   VARSPLIC    797    806       DCLRVGCAPS -> G (in some isoforms Alpha
FT                                3C).
FT                                /FTId=VSP_003513.
FT   VARSPLIC    807    809       Missing (in some isoforms Alpha 3B).
FT                                /FTId=VSP_003514.
FT   VARSPLIC   1256   1285       Missing (in some isoforms Alpha 4B).
FT                                /FTId=VSP_003515.
FT   VARSPLIC   1421   1614       Missing (in some isoforms Alpha 5B).
FT                                /FTId=VSP_003516.
FT   VARSPLIC   1667   1715       DEGSYQVDQSRNYISNSAQSNGAVVKEKAPAAPKTPSKAKK
FT                                NKDKEYYV -> CRKSPREEKLLPGSAQGLGLDLAKACCVC
FT                                RCRATCIAGKPLEERGGGRGEGERQMQIYIKNK (in
FT                                some isoforms Alpha 6).
FT                                /FTId=VSP_003517.
FT   VARIANT     434    434       N -> L.
SQ   SEQUENCE   1715 AA;  185282 MW;  59FBF18661F3DB15 CRC64;

  Query Match             16.9%;  Score 188;  DB 1;  Length 1715;
  Best Local Similarity   29.9%;  Pred. No. 2.2e-07;
  Matches   53;  Conservative   30;  Mismatches   56;  Indels   38;  Gaps    7;

Qy         28 DGFLAFPGHVFSRSLPEVPETIELEVRTSTASGLLLWQ-------GVEVGEAGQGKDFIS 80
              : |:| |     |:      :| |: ||:  :||||:        ||    : |  |: :
Db        501 EAFVALPRWSAKRT-----GSISLDFRTTEPNGLLLFSQGRRAGAGVGSHSSSQRADYFA 555

Qy         81 LGLQDGHLVFRYQLGSGEARL-VSEDPINDGEWHRVTALREGRRGSIQVDGEE---LVSG 136
              : | ||:|     :|||  :|  |   :|||||  |   |:||:||| |:      | :|
Db        556 MELLDGYLYLLLDMGSGGIKLRASSRKVNDGEWCHVDFQRDGRKGSISVNSRSTPFLATG 615

Qy        137 RSPGPNVAVNAKGSVYIGGAPDVATLTGGR-------------FSSGITGCVKNLVL 180
               |      :: :  :|:|| |:     |||               :|  |||::| :
Db        616 ESE----VLDLESELYLGGLPE-----GGRVDLPLPPEVWTAALRAGYVGCVRDLFI 663


RESULT 12
LML2_CAEEL
ID   LML2_CAEEL     STANDARD;      PRT;  3672 AA.
AC   Q21313;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Laminin-like protein K08C7.3 precursor.
GN   K08C7.3.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC   Rhabditidae; Peloderinae; Caenorhabditis.
OX   NCBI_TaxID=6239;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Bristol N2;
RA   Berks M.;
RL   Submitted (MAR-1996) to the EMBL/GenBank/DDBJ databases.
CC   -!- SIMILARITY: Contains 1 laminin N-terminal domain.
CC   -!- SIMILARITY: Contains 22 laminin EGF-like domains.
CC   -!- SIMILARITY: Contains 1 laminin IV domain.
CC   -!- SIMILARITY: Contains 5 laminin G-like domains.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
```

```
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; Z70286; CAA94293.1; -.
DR   PIR; T23433; T23433.
DR   HSSP; P02468; 1KLO.
DR   WormPep; K08C7.3; CE06136.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR006209; EGF_like.
DR   InterPro; IPR000034; Laminin_B.
DR   InterPro; IPR002049; Laminin_EGF.
DR   InterPro; IPR001791; Laminin_G.
DR   InterPro; IPR008211; LamNT.
DR   Pfam; PF00052; laminin_B; 1.
DR   Pfam; PF00053; laminin_EGF; 20.
DR   Pfam; PF00054; laminin_G; 5.
DR   Pfam; PF00055; laminin_Nterm; 1.
DR   PRINTS; PR00011; EGFLAMININ.
DR   SMART; SM00180; EGF_Lam; 17.
DR   SMART; SM00281; LamB; 1.
DR   SMART; SM00282; LamG; 5.
DR   SMART; SM00136; LamNT; 1.
DR   PROSITE; PS00022; EGF_1; 19.
DR   PROSITE; PS01186; EGF_2; 4.
DR   PROSITE; PS01248; LAMININ_TYPE_EGF; 21.
DR   PROSITE; PS50025; LAM_G_DOMAIN; 5.
KW   Hypothetical protein; Laminin EGF-like domain; Signal; Repeat.
FT   SIGNAL        1     27       POTENTIAL.
FT   CHAIN        28   3672       LAMININ-LIKE PROTEIN K08C7.3.
FT   DOMAIN       28    297       LAMININ N-TERMINAL (DOMAIN VI).
FT   DOMAIN      298    356       LAMININ EGF-LIKE 1.
FT   DOMAIN      357    426       LAMININ EGF-LIKE 2.
FT   DOMAIN      427    471       LAMININ EGF-LIKE 3.
FT   DOMAIN      472    518       LAMININ EGF-LIKE 4.
FT   DOMAIN      519    563       LAMININ EGF-LIKE 5.
FT   DOMAIN      564    609       LAMININ EGF-LIKE 6.
FT   DOMAIN      610    655       LAMININ EGF-LIKE 7.
FT   DOMAIN      656    700       LAMININ EGF-LIKE 8.
FT   DOMAIN      701    755       LAMININ EGF-LIKE 9.
FT   DOMAIN      756    808       LAMININ EGF-LIKE 10.
FT   DOMAIN      809    839       LAMININ EGF-LIKE 11 (INCOMPLETE).
FT   DOMAIN     1415   1460       LAMININ EGF-LIKE 12.
FT   DOMAIN     1461   1505       LAMININ EGF-LIKE 13.
FT   DOMAIN     1506   1553       LAMININ EGF-LIKE 14.
FT   DOMAIN     1554   1604       LAMININ EGF-LIKE 15.
FT   DOMAIN     1605   1614       LAMININ EGF-LIKE 16 (N-TERMINAL).
FT   DOMAIN     1615   1796       LAMININ DOMAIN IV.
FT   DOMAIN     1797   1829       LAMININ EGF-LIKE 16 (C-TERMINAL).
FT   DOMAIN     1830   1879       LAMININ EGF-LIKE 17.
FT   DOMAIN     1880   1936       LAMININ EGF-LIKE 18.
FT   DOMAIN     1937   1989       LAMININ EGF-LIKE 19.
FT   DOMAIN     1990   2036       LAMININ EGF-LIKE 20.
FT   DOMAIN     2037   2083       LAMININ EGF-LIKE 21.
FT   DOMAIN     2084   2131       LAMININ EGF-LIKE 22.
FT   DOMAIN     2693   2884       LAMININ G-LIKE 1.
FT   DOMAIN     2896   3066       LAMININ G-LIKE 2.
FT   DOMAIN     3072   3235       LAMININ G-LIKE 3.
FT   DOMAIN     3310   3482       LAMININ G-LIKE 4.
FT   DOMAIN     3488   3669       LAMININ G-LIKE 5.
FT   DISULFID    298    307       BY SIMILARITY.
FT   DISULFID    300    320       BY SIMILARITY.
FT   DISULFID    322    331       BY SIMILARITY.
FT   DISULFID    334    354       BY SIMILARITY.
FT   DISULFID    357    366       BY SIMILARITY.
FT   DISULFID    359    391       BY SIMILARITY.
FT   DISULFID    394    403       BY SIMILARITY.
FT   DISULFID    406    424       BY SIMILARITY.
FT   DISULFID    427    438       BY SIMILARITY.
FT   DISULFID    429    445       BY SIMILARITY.
FT   DISULFID    447    456       BY SIMILARITY.
FT   DISULFID    459    469       BY SIMILARITY.
FT   DISULFID    472    484       BY SIMILARITY.
FT   DISULFID    474    491       BY SIMILARITY.
```

```
FT   DISULFID    493    502       BY SIMILARITY.
FT   DISULFID    505    516       BY SIMILARITY.
FT   DISULFID    519    531       BY SIMILARITY.
FT   DISULFID    521    538       BY SIMILARITY.
FT   DISULFID    540    549       BY SIMILARITY.
FT   DISULFID    552    561       BY SIMILARITY.
FT   DISULFID    564    576       BY SIMILARITY.
FT   DISULFID    566    583       BY SIMILARITY.
FT   DISULFID    585    594       BY SIMILARITY.
FT   DISULFID    597    607       BY SIMILARITY.
FT   DISULFID    610    622       BY SIMILARITY.
FT   DISULFID    612    629       BY SIMILARITY.
FT   DISULFID    631    640       BY SIMILARITY.
FT   DISULFID    643    653       BY SIMILARITY.
FT   DISULFID    656    668       BY SIMILARITY.
FT   DISULFID    658    674       BY SIMILARITY.
FT   DISULFID    676    685       BY SIMILARITY.
FT   DISULFID    688    698       BY SIMILARITY.
FT   DISULFID    701    715       BY SIMILARITY.
FT   DISULFID    703    724       BY SIMILARITY.
FT   DISULFID    726    735       BY SIMILARITY.
FT   DISULFID    738    753       BY SIMILARITY.
FT   DISULFID    756    770       BY SIMILARITY.
FT   DISULFID    758    777       BY SIMILARITY.
FT   DISULFID    779    788       BY SIMILARITY.
FT   DISULFID    791    806       BY SIMILARITY.
FT   DISULFID   1415   1427       BY SIMILARITY.
FT   DISULFID   1417   1434       BY SIMILARITY.
FT   DISULFID   1436   1445       BY SIMILARITY.
FT   DISULFID   1448   1458       BY SIMILARITY.
FT   DISULFID   1461   1469       BY SIMILARITY.
FT   DISULFID   1463   1476       BY SIMILARITY.
FT   DISULFID   1478   1487       BY SIMILARITY.
FT   DISULFID   1490   1503       BY SIMILARITY.
FT   DISULFID   1506   1520       BY SIMILARITY.
FT   DISULFID   1508   1527       BY SIMILARITY.
FT   DISULFID   1529   1538       BY SIMILARITY.
FT   DISULFID   1541   1551       BY SIMILARITY.
FT   DISULFID   1554   1566       BY SIMILARITY.
FT   DISULFID   1556   1573       BY SIMILARITY.
FT   DISULFID   1575   1584       BY SIMILARITY.
FT   DISULFID   1587   1602       BY SIMILARITY.
FT   DISULFID   1830   1839       BY SIMILARITY.
FT   DISULFID   1832   1846       BY SIMILARITY.
FT   DISULFID   1849   1858       BY SIMILARITY.
FT   DISULFID   1861   1877       BY SIMILARITY.
FT   DISULFID   1880   1894       BY SIMILARITY.
FT   DISULFID   1882   1905       BY SIMILARITY.
FT   DISULFID   1907   1916       BY SIMILARITY.
FT   DISULFID   1919   1934       BY SIMILARITY.
FT   DISULFID   1937   1951       BY SIMILARITY.
FT   DISULFID   1939   1958       BY SIMILARITY.
FT   DISULFID   1961   1970       BY SIMILARITY.
FT   DISULFID   1973   1987       BY SIMILARITY.
FT   DISULFID   1990   2000       BY SIMILARITY.
FT   DISULFID   1992   2007       BY SIMILARITY.
FT   DISULFID   2009   2018       BY SIMILARITY.
FT   DISULFID   2021   2034       BY SIMILARITY.
FT   DISULFID   2037   2048       BY SIMILARITY.
FT   DISULFID   2039   2055       BY SIMILARITY.
FT   DISULFID   2057   2066       BY SIMILARITY.
FT   DISULFID   2069   2081       BY SIMILARITY.
FT   DISULFID   2084   2096       BY SIMILARITY.
FT   DISULFID   2086   2103       BY SIMILARITY.
FT   DISULFID   2105   2114       BY SIMILARITY.
FT   DISULFID   2117   2129       BY SIMILARITY.
FT   CARBOHYD    121    121       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    140    140       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    249    249       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    351    351       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    477    477       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    511    511       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    530    530       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    634    634       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    761    761       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1014   1014       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1341   1341       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1705   1705       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1756   1756       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1868   1868       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1944   1944       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1986   1986       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   2002   2002       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   2159   2159       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   2207   2207       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   2231   2231       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   2235   2235       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   2401   2401       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   2421   2421       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   2487   2487       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   2821   2821       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   3087   3087       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   3242   3242       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   3541   3541       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   3672 AA;  404223 MW;  28E262DB5FF14BFA CRC64;

  Query Match             16.6%;  Score 185.5;  DB 1;  Length 3672;
  Best Local Similarity   29.0%;  Pred. No. 8.5e-07;
  Matches   54;  Conservative   30;  Mismatches   55;  Indels   47;  Gaps    8;

Qy          28 DGFLAF-PGHVFSRSLPEVPETIELEVRTSTASGLLLWQGVEVGEAGQGKDFISLGLQDG 86
               :|: :: | |   |   :    | |   | :  ||| :        |: |||::| | ||
Db        2904 EGYTSYKPSHWNPRKATK----ISLSFLTFSPHGLLFF-------VGKDKDFMALELSDG 2952

Qy          87 HLVFRYQLGSGEARLVSEDP-INDGEWHRVTALREGRRGSIQVDGE-ELVSGRSPGPNVA 144
                :     ||||  : ::|    |||:|| |: :|| :   | :||| |:: |  || :
Db        2953 GVKLSVDLGSGVGQWITESSNYNDGKWHTVSIVREEKHVKIMIDGETEVLEGDVPGKDSE 3012

Qy         145 VNAKGSVYIGGAPDVATLTGGRFSSGIT---------GCVKNLVLHS------------- 182
               ::    :|||| |           ||::         ||:|:: | |
Db        3013 MSVTEFLYIGGTP-----------SGLSVRTTIVPLRGCIKSVKLGSDNVDLESSHASKG 3061

Qy         183 ARPGAP 188
                | | |
Db        3062 VRSGCP 3067


RESULT 13
LMA4_MOUSE
ID   LMA4_MOUSE     STANDARD;      PRT;  1816 AA.
AC   P97927; O88785; P70409;
DT   28-FEB-2003 (Rel. 41, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   15-MAR-2004 (Rel. 43, Last annotation update)
DE   Laminin alpha-4 chain precursor.
GN   LAMA4.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A., AND SEQUENCE OF 462-469; 478-483; 776-782 AND
RP   940-945.
RC   STRAIN=BALB/c; TISSUE=Endothelial cells;
RX   MEDLINE=97363207; PubMed=9219532;
RA   Frieser M., Noeckel H., Pausch F., Roeder C., Hahn A., Deutzmann R.,
RA   Sorokin L.M.;
RT   "Cloning of the mouse laminin alpha 4 cDNA. Expression in a subset of
RT   endothelium.";
RL   Eur. J. Biochem. 246:727-735(1997).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Heart;
RX   MEDLINE=97202462; PubMed=9049981;
RA   Liu J., Mayne R.;
RT   "The complete cDNA coding sequence and tissue-specific expression of
```

```
RT   the mouse laminin alpha 4 chain.";
RL   Matrix Biol. 15:433-437(1996).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BALB/c;
RX   MEDLINE=98010627; PubMed=9346933;
RA   Iivanainen A., Kortesmaa J., Sahlberg C., Morita T., Bergmann U.,
RA   Thesleff I., Tryggvason K.;
RT   "Primary structure, developmental expression, and immunolocalization
RT   of the murine laminin alpha4 chain.";
RL   J. Biol. Chem. 272:27862-27868(1997).
RN   [4]
RP   SEQUENCE OF 836-1106 FROM N.A.
RC   STRAIN=ICR; TISSUE=Placenta;
RX   MEDLINE=97296337; PubMed=9151674;
RA   Miner J.H., Patton B.L., Lentz S.I., Gilbert D.J., Snider W.D.,
RA   Jenkins N.A., Copeland N.G., Sanes J.R.;
RT   "The laminin alpha chains: expression, developmental transitions, and
RT   chromosomal locations of alpha1-5, identification of heterotrimeric
RT   laminins 8-11, and cloning of a novel alpha3 isoform.";
RL   J. Cell Biol. 137:685-702(1997).
RN   [5]
RP   SEQUENCE OF 1467-1691 FROM N.A.
RC   TISSUE=Placenta;
RX   MEDLINE=97187457; PubMed=9034910;
RA   Lentz S.I., Miner J.H., Sanes J.R., Snider W.D.;
RT   "Distribution of the ten known laminin chains in the pathways and
RT   targets of developing sensory axons.";
RL   J. Comp. Neurol. 378:547-561(1997).
CC   -!- FUNCTION: Binding to cells via a high affinity receptor, laminin
CC       is thought to mediate the attachment, migration and organization
CC       of cells into tissues during embryonic development by interacting
CC       with other extracellular matrix components.
CC   -!- SUBUNIT: Laminin is a complex glycoprotein, consisting of three
CC       different polypeptide chains (alpha, beta, gamma), which are bound
CC       to each other by disulfide bonds into a cross-shaped molecule
CC       comprising one long and three short arms with globules at each
CC       end.
CC   -!- SUBCELLULAR LOCATION: Extracellular; found in the basement
CC       membranes (major component).
CC   -!- TISSUE SPECIFICITY: Strongly expressed in peripheral nerves,
CC       cardiac muscle, fat, dermis, lung stroma, aortic endothelium,
CC       endocardium and endothelium of blood vessels in skin and brain.
CC   -!- DOMAIN: The alpha-helical domains I and II are thought to interact
CC       with other laminin chains to form a coiled coil structure.
CC   -!- DOMAIN: Domain G is globular.
CC   -!- SIMILARITY: Contains 4 laminin EGF-like domains.
CC   -!- SIMILARITY: Contains 5 laminin G-like domains.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; U58950; AAB41840.1; -.
DR   EMBL; Y09827; CAA70970.1; -.
DR   EMBL; U59865; AAC24725.1; -.
DR   EMBL; U88352; AAC53178.1; -.
DR   EMBL; U69176; AAC52982.1; -.
DR   HSSP; P02468; 1KLO.
DR   MGD; MGI:109321; Lama4.
DR   GO; GO:0005604; C:basement membrane; IDA.
DR   InterPro; IPR006209; EGF_like.
DR   InterPro; IPR002049; Laminin_EGF.
DR   InterPro; IPR001791; Laminin_G.
DR   Pfam; PF00053; laminin_EGF; 3.
DR   Pfam; PF00054; laminin_G; 4.
DR   SMART; SM00180; EGF_Lam; 3.
DR   SMART; SM00282; LamG; 5.
DR   PROSITE; PS00022; EGF_1; UNKNOWN_1.
```

```
DR   PROSITE; PS01248; LAMININ_TYPE_EGF; 3.
DR   PROSITE; PS50025; LAM_G_DOMAIN; 5.
KW   Glycoprotein; Basement membrane; Extracellular matrix; Coiled coil;
KW   Laminin EGF-like domain; Cell adhesion; Repeat; Signal.
FT   SIGNAL        1     24       POTENTIAL.
FT   CHAIN        25   1816       LAMININ ALPHA-4 CHAIN.
FT   DOMAIN       82    131       LAMININ EGF-LIKE 1.
FT   DOMAIN      132    186       LAMININ EGF-LIKE 2.
FT   DOMAIN      187    240       LAMININ EGF-LIKE 3.
FT   DOMAIN      241    255       LAMININ EGF-LIKE 4 (INCOMPLETE).
FT   DOMAIN      256    825       DOMAIN II AND I.
FT   DOMAIN      826   1030       LAMININ G-LIKE 1.
FT   DOMAIN     1042   1222       LAMININ G-LIKE 2.
FT   DOMAIN     1229   1397       LAMININ G-LIKE 3.
FT   DOMAIN     1462   1633       LAMININ G-LIKE 4.
FT   DOMAIN     1640   1813       LAMININ G-LIKE 5.
FT   DOMAIN      431    523       COILED COIL (POTENTIAL).
FT   DOMAIN      556    604       COILED COIL (POTENTIAL).
FT   DOMAIN      655    717       COILED COIL (POTENTIAL).
FT   DOMAIN      770    799       COILED COIL (POTENTIAL).
FT   SITE        717    719       CELL ATTACHMENT SITE (POTENTIAL).
FT   DISULFID     82     91       BY SIMILARITY.
FT   DISULFID     84     98       BY SIMILARITY.
FT   DISULFID    101    110       BY SIMILARITY.
FT   DISULFID    113    129       BY SIMILARITY.
FT   DISULFID    132    146       BY SIMILARITY.
FT   DISULFID    134    155       BY SIMILARITY.
FT   DISULFID    157    166       BY SIMILARITY.
FT   DISULFID    169    184       BY SIMILARITY.
FT   DISULFID    187    202       BY SIMILARITY.
FT   DISULFID    189    209       BY SIMILARITY.
FT   DISULFID    212    221       BY SIMILARITY.
FT   DISULFID    224    238       BY SIMILARITY.
FT   DISULFID    266    266       INTERCHAIN (PROBABLE).
FT   DISULFID    269    269       INTERCHAIN (PROBABLE).
FT   CARBOHYD    104    104       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    215    215       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    308    308       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    333    333       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    458    458       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    550    550       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    571    571       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    574    574       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    631    631       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    639    639       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    735    735       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    751    751       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    754    754       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    780    780       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    803    803       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1088   1088       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1283   1283       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1361   1361       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CONFLICT      8      8       C -> S (IN REF. 2).
FT   CONFLICT     18     18       C -> Y (IN REF. 2).
FT   CONFLICT    248    248       C -> R (IN REF. 3).
FT   CONFLICT    297    297       G -> A (IN REF. 3).
FT   CONFLICT    431    433       THR -> HPS (IN REF. 2).
FT   CONFLICT    679    679       S -> C (IN REF. 3).
FT   CONFLICT    703    703       D -> G (IN REF. 2).
FT   CONFLICT    706    706       N -> H (IN REF. 2).
FT   CONFLICT    728    728       K -> R (IN REF. 2).
FT   CONFLICT    730    730       F -> I (IN REF. 2).
FT   CONFLICT    779    779       R -> G (IN REF. 1; AA SEQUENCE).
FT   CONFLICT    810    810       R -> S (IN REF. 3).
FT   CONFLICT    865    867       AEP -> QT (IN REF. 2).
FT   CONFLICT    936    936       K -> E (IN REF. 3).
FT   CONFLICT    970    970       L -> V (IN REF. 3).
FT   CONFLICT   1132   1132       H -> R (IN REF. 2).
FT   CONFLICT   1200   1200       F -> I (IN REF. 2).
FT   CONFLICT   1382   1382       D -> A (IN REF. 2).
FT   CONFLICT   1413   1414       NS -> EF (IN REF. 1).
FT   CONFLICT   1489   1489       A -> S (IN REF. 2).
```

```
SQ   SEQUENCE   1816 AA;  201818 MW;  B49C45F3A45999D8 CRC64;

  Query Match             15.6%;  Score 173.5;  DB 1;  Length 1816;
  Best Local Similarity   28.1%;  Pred. No. 3.5e-06;
  Matches   38;  Conservative   32;  Mismatches   58;  Indels    7;  Gaps    2;

Qy           51 LEVRTSTASGLLLWQGVEVGEAGQGKDFISLGLQDGHLVFRYQLGSGEARLVSEDPINDG 110
                : ::| :: |:: :   :     :  ||::| |  | ||| : :|  : :: |::  |||
Db         1490 IRLKTRSSHGMIFYVSDQ-----EENDFMTLFLAHGRLVFMFNVGHKKLKIRSQEKYNDG 1544

Qy          111 EWHRVTALREGRRGSIQVDGEELVSGRSPGPNVAVNAKGSVYIGGAPDVATLTGGRFSS- 169
                 || |  :||   | : :||  ::  | |    |   || :|:||       :   : :|
Db         1545 LWHDVIFIREKSSGRLVIDGLRVLEERLPPSGAAWKIKGPIYLGGVAPGRAVKNVQITSV 1604

Qy          170 -GITGCVKNLVLHSA 183
                   :||: || |: |
Db         1605 YSFSGCLGNLQLNGA 1619


RESULT 14
LMA3_MOUSE
ID   LMA3_MOUSE     STANDARD;      PRT;  3333 AA.
AC   Q61789; O08751; Q61788; Q61966; Q9JHQ7;
DT   01-NOV-1997 (Rel. 35, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   15-MAR-2004 (Rel. 43, Last annotation update)
DE   Laminin alpha-3 chain precursor (Nicein alpha subunit).
GN   LAMA3.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE OF 1-58 FROM N.A., AND SEQUENCE OF 32-38.
RX   MEDLINE=21818471; PubMed=11829758;
RA   Garbe J.H., Gohring W., Mann K., Timpl R., Sasaki T.;
RT   "Complete sequence, recombinant analysis and binding to laminins and
RT   sulphated ligands of the N-terminal domains of laminin alpha3B and
RT   alpha5 chains.";
RL   Biochem. J. 362:213-221(2002).
RN   [2]
RP   SEQUENCE OF 1-726 FROM N.A.
RC   STRAIN=ICR;
RX   MEDLINE=97296337; PubMed=9151674;
RA   Miner J.H., Patton B.L., Lentz S.I., Gilbert D.J., Snider W.D.,
RA   Jenkins N.A., Copeland N.G., Sanes J.R.;
RT   "The laminin alpha chains: expression, developmental transitions, and
RT   chromosomal locations of alpha1-5, identification of heterotrimeric
RT   laminins 8-11, and cloning of a novel alpha3 isoform.";
RL   J. Cell Biol. 137:685-701(1997).
RN   [3]
RP   SEQUENCE FROM N.A. (ISOFORM A), AND SEQUENCE OF 716-3284 FROM N.A.
RP   (ISOFORM B).
RC   STRAIN=BALB/c; TISSUE=Lung;
RX   MEDLINE=95394948; PubMed=7665604;
RA   Galliano M.-F., Aberdam D., Aguzzi A., Ortonne J.-P., Meneguzzi G.;
RT   "Cloning and complete primary structure of the mouse laminin alpha 3
RT   chain. Distinct expression pattern of the laminin alpha 3A and alpha
RT   3B chain isoforms.";
RL   J. Biol. Chem. 270:21820-21826(1995).
RN   [4]
RP   REVISIONS.
RA   Aberdam D.;
RL   Submitted (APR-1997) to the EMBL/GenBank/DDBJ databases.
RN   [5]
RP   SEQUENCE OF 1767-2485 FROM N.A.
RC   TISSUE=Lung;
RX   MEDLINE=94281750; PubMed=8012114;
RA   Aberdam D., Galliano M.-F., Mattei M.-G., Pisani-Spadafora A.,
RA   Ortonne J.-P., Meneguzzi G.;
RT   "Assignment of mouse nicein genes to chromosomes 1 and 18.";
RL   Mamm. Genome 5:229-233(1994).
RN   [6]
RP   SEQUENCE OF 1767-2485 FROM N.A.
RC   TISSUE=Lung;
RX   MEDLINE=94363405; PubMed=8081888;
RA   Aberdam D., Aguzzi A., Baudoin C., Galliano M.-F., Ortonne J.-P.,
RA   Meneguzzi G.;
RT   "Developmental expression of nicein adhesion protein (laminin-5)
RT   subunits suggests multiple morphogenic roles.";
RL   Cell Adhes. Commun. 2:115-129(1994).
CC   -!- FUNCTION: Binding to cells via a high affinity receptor, laminin
CC       is thought to mediate the attachment, migration and organization
CC       of cells into tissues during embryonic development by interacting
CC       with other extracellular matrix components.
CC   -!- FUNCTION: Laminin-5 is thought to be involved in (1) cell adhesion
CC       via integrin alpha-3/beta-1 in focal adhesion and integrin
CC       alpha-6/beta-4 in hemidesmosomes, (2) signal transduction via
CC       tyrosine phosphorylation of pp125-FAK and p80, (3) differentiation
CC       of keratinocytes (By similarity).
CC   -!- SUBUNIT: Laminin is a complex glycoprotein, consisting of three
CC       different polypeptide chains (alpha, beta, gamma), which are bound
CC       to each other by disulfide bonds into a cross-shaped molecule
CC       comprising one long and three short arms with globules at each
CC       end. The alpha-3 chain is a subunit of laminin-5
CC       (epiligrin/kalinin/nicein), and possibly also a component of
CC       laminin-6 (K-laminin) and laminin-7 (KS-laminin).
CC   -!- SUBCELLULAR LOCATION: Extracellular; found in the basement
CC       membranes (major component).
CC   -!- ALTERNATIVE PRODUCTS:
CC       Event=Alternative splicing; Named isoforms=2;
CC       Name=B;
CC         IsoId=Q61789-1; Sequence=Displayed;
CC       Name=A;
CC         IsoId=Q61789-2; Sequence=VSP_003038, VSP_003039;
CC   -!- TISSUE SPECIFICITY: Basal membrane of the upper alimentary tract
CC       and urinary and nasal epithelia, salivary glands and teeth (both
CC       variants). Isoform A is predominantly expressed in skin, hair
CC       follicles and developing neurons of the trigeminal ganglion.
CC       Isoform B was found in bronchi, alveoli, stomach, intestinal
CC       crypts, whisker pads, CNS, telencephalic neuroectoderm, thalamus,
CC       Rathke's pouch and periventricular subependymal germinal layer.
CC   -!- DOMAIN: The alpha-helical domains I and II are thought to interact
CC       with other laminin chains to form a coiled coil structure.
CC   -!- DOMAIN: Domains IV and G are globular.
CC   -!- SIMILARITY: Contains 1 laminin N-terminal domain.
CC   -!- SIMILARITY: Contains 15 laminin EGF-like domains.
CC   -!- SIMILARITY: Contains 1 laminin IV domain.
CC   -!- SIMILARITY: Contains 5 laminin G-like domains.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; AJ293592; CAB99254.2; -.
DR   EMBL; U88353; AAC53179.1; -.
DR   EMBL; X84014; CAA58837.1; -.
DR   EMBL; X84013; CAA58836.1; ALT_FRAME.
DR   EMBL; L20478; AAA68091.1; -.
DR   HSSP; P02468; 1TLE.
DR   MGD; MGI:99909; Lama3.
DR   GO; GO:0005604; C:basement membrane; IDA.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR006209; EGF_like.
DR   InterPro; IPR008212; Lam_N2.
DR   InterPro; IPR000034; Laminin_B.
DR   InterPro; IPR002049; Laminin_EGF.
DR   InterPro; IPR001791; Laminin_G.
DR   InterPro; IPR008211; LamNT.
DR   Pfam; PF00052; laminin_B; 1.
DR   Pfam; PF00053; laminin_EGF; 10.
DR   Pfam; PF00054; laminin_G; 3.
```

```
DR   Pfam; PF00055; laminin_Nterm; 1.
DR   PRINTS; PR00011; EGFLAMININ.
DR   ProDom; PD002082; Lam_N2; 1.
DR   ProDom; PD003031; Laminin_B; 1.
DR   SMART; SM00180; EGF_Lam; 13.
DR   SMART; SM00281; LamB; 1.
DR   SMART; SM00282; LamG; 5.
DR   SMART; SM00136; LamNT; 1.
DR   PROSITE; PS00022; EGF_1; 10.
DR   PROSITE; PS01186; EGF_2; 1.
DR   PROSITE; PS01248; LAMININ_TYPE_EGF; 10.
DR   PROSITE; PS50025; LAM_G_DOMAIN; 5.
KW   Glycoprotein; Basement membrane; Extracellular matrix; Coiled coil;
KW   Laminin EGF-like domain; Cell adhesion; Repeat; Signal;
KW   Alternative splicing.
FT   SIGNAL        1     31
FT   CHAIN        32   3333       LAMININ ALPHA-3 CHAIN.
FT   DOMAIN       32    294       LAMININ N-TERMINAL (DOMAIN VI).
FT   DOMAIN      295    725       DOMAIN V.
FT   DOMAIN      296    350       LAMININ EGF-LIKE 1.
FT   DOMAIN      353    420       LAMININ EGF-LIKE 2.
FT   DOMAIN      423    464       LAMININ EGF-LIKE 3.
FT   DOMAIN      488    530       LAMININ EGF-LIKE 4.
FT   DOMAIN      533    576       LAMININ EGF-LIKE 5.
FT   DOMAIN      582    625       LAMININ EGF-LIKE 6.
FT   DOMAIN      628    678       LAMININ EGF-LIKE 7.
FT   DOMAIN      681    725       LAMININ EGF-LIKE 8.
FT   DOMAIN      793   1262       LAMININ DOMAIN IV 1 (DOMAIN IV B).
FT   DOMAIN     1263   1464       DOMAIN III B.
FT   DOMAIN     1310   1353       LAMININ EGF-LIKE 9.
FT   DOMAIN     1354   1403       LAMININ EGF-LIKE 10.
FT   DOMAIN     1404   1454       LAMININ EGF-LIKE 11.
FT   DOMAIN     1455   1464       LAMININ EGF-LIKE 12 (N-TERMINAL).
FT   DOMAIN     1465   1653       LAMININ DOMAIN IV (DOMAIN IV A).
FT   DOMAIN     1654   1821       DOMAIN III A.
FT   DOMAIN     1654   1686       LAMININ EGF-LIKE 12 (C-TERMINAL).
FT   DOMAIN     1687   1733       LAMININ EGF-LIKE 13.
FT   DOMAIN     1734   1786       LAMININ EGF-LIKE 14.
FT   DOMAIN     1787   1821       LAMININ EGF-LIKE 15 (INCOMPLETE).
FT   DOMAIN     1822   2388       DOMAIN II AND I.
FT   DOMAIN     2389   2590       LAMININ G-LIKE 1.
FT   DOMAIN     2597   2759       LAMININ G-LIKE 2.
FT   DOMAIN     2766   2926       LAMININ G-LIKE 3.
FT   DOMAIN     2986   3150       LAMININ G-LIKE 4.
FT   DOMAIN     3157   3330       LAMININ G-LIKE 5.
FT   DOMAIN     1854   1983       COILED COIL (POTENTIAL).
FT   DOMAIN     2015   2060       COILED COIL (POTENTIAL).
FT   DOMAIN     2091   2168       COILED COIL (POTENTIAL).
FT   DOMAIN     2214   2241       COILED COIL (POTENTIAL).
FT   DOMAIN     2321   2386       COILED COIL (POTENTIAL).
FT   SITE       2277   2279       CELL ATTACHMENT SITE (POTENTIAL).
FT   DISULFID   1310   1317       BY SIMILARITY.
FT   DISULFID   1312   1324       BY SIMILARITY.
FT   DISULFID   1326   1335       BY SIMILARITY.
FT   DISULFID   1338   1351       BY SIMILARITY.
FT   DISULFID   1354   1369       BY SIMILARITY.
FT   DISULFID   1356   1376       BY SIMILARITY.
FT   DISULFID   1378   1387       BY SIMILARITY.
FT   DISULFID   1390   1401       BY SIMILARITY.
FT   DISULFID   1404   1416       BY SIMILARITY.
FT   DISULFID   1406   1423       BY SIMILARITY.
FT   DISULFID   1425   1434       BY SIMILARITY.
FT   DISULFID   1437   1452       BY SIMILARITY.
FT   DISULFID   1687   1696       BY SIMILARITY.
FT   DISULFID   1689   1703       BY SIMILARITY.
FT   DISULFID   1706   1715       BY SIMILARITY.
FT   DISULFID   1718   1731       BY SIMILARITY.
FT   DISULFID   1734   1746       BY SIMILARITY.
FT   DISULFID   1736   1755       BY SIMILARITY.
FT   DISULFID   1757   1766       BY SIMILARITY.
FT   DISULFID   1769   1784       BY SIMILARITY.
FT   DISULFID   1822   1822       INTERCHAIN (PROBABLE).
FT   DISULFID   1825   1825       INTERCHAIN (PROBABLE).
FT   CARBOHYD    139    139       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    445    445       N-LINKED (GLCNAC. . .) (POTENTIAL).

  Query Match             15.3%;  Score 171;  DB 1;  Length 3333;
  Best Local Similarity   27.6%;  Pred. No. 1.2e-05;
  Matches   51;  Conservative   34;  Mismatches   72;  Indels   28;  Gaps    6;

Qy          31 LAFPGHVFSRSLPEVPETI-------ELEVRTSTASGLLLWQGVEVGEAGQGKDFISLGL 83
               | |     |  | ::|: :        |:::|::: ||       |   |    |::| |
Db        2987 LQFGDSPTSHLLFKLPQELLKPRLQFSLDIQTTSSRGL-------VFHTGTRDSFVALYL 3039

Qy          84 QDGHLVFRYQLGSGEARLVSEDPINDGEWHRVTALREGRRGSIQVDGEELVSGRSPGPNV 143
                :||::|    |  : || |::  :||:|| |     ||:  : |||     |  || |
Db        3040 SEGHVIFALGAGGKKLRLRSKERYHDGKWHSVVFGLSGRKVHLVVDGLRAQEGSLPG-NS 3098

Qy         144 AVNAKGSVYIGGAPDVATLTGGRFSSGITGCVKNLVLHSARPGAPPPQPLDLQHRAQAGA 203
                :: :  ||:| :|   : :  : |    ||::|  | |        :|||      | :
Db        3099 TISPREQVYLGLSPSRKSKSLPQHS--FVGCLRNFQLDS--------KPLD---SPSARS 3145

Qy         204 NTRPC 208
                  ||
Db        3146 GVSPC 3150


RESULT 15
NX3A_HUMAN
ID   NX3A_HUMAN     STANDARD;      PRT;  1541 AA.
AC   Q9Y4C0; O95378; Q9NS47; Q9P1V3; Q9P1V6; Q9UIE2; Q9UIE3; Q9ULA5;
AC   Q9Y486;
DT   28-FEB-2003 (Rel. 41, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   15-MAR-2004 (Rel. 43, Last annotation update)
DE   Neurexin 3-alpha precursor (Neurexin III-alpha).
GN   NRXN3 OR KIAA0743.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=21945268; PubMed=11944992;
RA   Rowen L., Young J., Birditt B., Kaur A., Madan A., Philipps D.L.,
RA   Qin S., Minx P., Wilson R.K., Hood L., Graveley B.R.;
RT   "Analysis of the human neurexin genes: alternative splicing and the
RT   generation of protein diversity.";
RL   Genomics 79:587-597(2002).
RN   [2]
RP   SEQUENCE OF 253-1541 FROM N.A.
RC   TISSUE=Brain;
RX   MEDLINE=99087487; PubMed=9872452;
RA   Nagase T., Ishikawa K.-I., Suyama M., Kikuno R., Miyajima N.,
RA   Tanaka A., Kotani H., Nomura N., Ohara O.;
RT   "Prediction of the coding sequences of unidentified human genes. XI.
RT   The complete sequences of 100 new cDNA clones from brain which code
RT   for large proteins in vitro.";
RL   DNA Res. 5:277-286(1998).
CC   -!- FUNCTION: Neuronal cell surface protein that may be involved in
CC       cell recognition and cell adhesion. May mediate intracellular
CC       signaling.
CC   -!- SUBUNIT: The laminin G-like domain 2 binds to neurexophilin 1.
CC       Specific isoforms bind to alpha-dystroglycan. The cytoplasmic C-
CC       terminal region binds to CASK (By similarity).
CC   -!- SUBCELLULAR LOCATION: Type I membrane protein (Potential).
CC   -!- ALTERNATIVE PRODUCTS:
CC       Event=Alternative promoter;
CC         Comment=A number of isoforms, alpha-type (shown here) and
CC         beta-type (AC Q9HDB5), are produced by use of alternative
CC         promoters. Beta-type isoforms differ from alpha-type isoforms
CC         in their N-terminus;
CC       Event=Alternative splicing; Named isoforms=1;
CC         Comment=A number of isoforms may be produced;
CC       Name=1;
CC         IsoId=Q9Y4C0-1; Sequence=Displayed;
```

```
CC   -!- TISSUE SPECIFICITY: Predominantly expressed in brain.
CC   -!- SIMILARITY: Contains 6 laminin G-like domains.
CC   -!- SIMILARITY: Contains 3 EGF-like domains.
CC   -!- SIMILARITY: Belongs to the neurexin family.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; AF099810; AAC68909.1; -.
DR   EMBL; AC008045; AAF28465.1; -.
DR   EMBL; AC009396; AAF21147.1; -.
DR   EMBL; AC012099; AAF15058.1; -.
DR   EMBL; AC026888; AAF87841.1; -.
DR   EMBL; AC011440; AAF61277.1; -.
DR   EMBL; AC008056; AAF09143.1; -.
DR   EMBL; AF123462; AAD13621.1; -.
DR   EMBL; AB018286; BAA34463.1; ALT_INIT.
DR   HSSP; Q63373; 1C4R.
DR   Genew; HGNC:8010; NRXN3.
DR   MIM; 600567; -.
DR   GO; GO:0005887; C:integral to plasma membrane; TAS.
DR   GO; GO:0004872; F:receptor activity; TAS.
DR   GO; GO:0007411; P:axon guidance; TAS.
DR   InterPro; IPR000152; Asx_hydroxyl_S.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR006209; EGF_like.
DR   InterPro; IPR001791; Laminin_G.
DR   Pfam; PF00008; EGF; 3.
DR   Pfam; PF00054; laminin_G; 6.
DR   PROSITE; PS00010; ASX_HYDROXYL; 1.
DR   PROSITE; PS00022; EGF_1; FALSE_NEG.
DR   PROSITE; PS01186; EGF_2; FALSE_NEG.
DR   PROSITE; PS50026; EGF_3; 3.
DR   PROSITE; PS50025; LAM_G_DOMAIN; 6.
KW   Signal; Transmembrane; Repeat; Cell adhesion; EGF-like domain;
KW   Glycoprotein; Alternative splicing; Alternative promoter usage.
FT   SIGNAL        1     27       BY SIMILARITY.
FT   CHAIN        28   1541       NEUREXIN 3-ALPHA.
FT   DOMAIN       28   1466       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   1467   1487       POTENTIAL.
FT   DOMAIN     1488   1541       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN       28    202       LAMININ G-LIKE 1.
FT   DOMAIN      198    235       EGF-LIKE 1.
FT   DOMAIN      258    440       LAMININ G-LIKE 2.
FT   DOMAIN      447    639       LAMININ G-LIKE 3.
FT   DOMAIN      643    680       EGF-LIKE 2.
FT   DOMAIN      685    857       LAMININ G-LIKE 4.
FT   DOMAIN      871   1046       LAMININ G-LIKE 5.
FT   DOMAIN     1049   1086       EGF-LIKE 3.
FT   DOMAIN     1090   1260       LAMININ G-LIKE 6.
FT   DOMAIN     1289   1292       POLY-THR.
FT   DOMAIN     1295   1298       POLY-THR.
FT   DOMAIN     1370   1373       POLY-THR.
FT   DOMAIN     1385   1388       POLY-SER.
FT   DOMAIN     1474   1477       POLY-ALA.
FT   CARBOHYD     58     58       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    105    105       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    757    757       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1189   1189       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1257   1257       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1301   1301       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CONFLICT   1043   1043       K -> E (IN REF. 1; AAF87841/AAF61277).
FT   CONFLICT   1334   1334       T -> TTGGELVIPLLVEDPLATPPIATRAPSITLPPTFRP
FT                                LLTIIETTKDSLSMTSEAGLPCLSDQGSDGCDDDGLVISGY
FT                                GSGETFDSNLPPTDDEDFYTTFSLVT (IN REF. 1;
FT                                AAD13621).
SQ   SEQUENCE   1541 AA;  169923 MW;  BE67FE2FE6197C95 CRC64;
```

```
  Query Match             15.1%;  Score 168;  DB 1;  Length 1541;
  Best Local Similarity   30.7%;  Pred. No. 8.2e-06;
  Matches   42;  Conservative   23;  Mismatches   60;  Indels   12;  Gaps    3;

Qy          47 ETIELEVRTSTASGLLLWQGVEVGEAGQGKDFISLGLQDGHLVFRYQLGSGEARLVSEDP 106
               : | |  :|   :||:|         |:  |:::| |:|| :     ||||    : | |
Db         281 DEITLSFKTWQRNGLIL-------HTGKSADYVNLALKDGAVSLVINLGSGAFEAIVE-P 332

Qy         107 I----NDGEWHRVTALREGRRGSIQVDGEELVSGRSPGPNVAVNAKGSVYIGGAPDVATL 162
               :    ||  || |   |  |: :| |||    :| :      : :    |:||:|  | |
Db         333 VNGKFNDNAWHDVKVTRNLRQVTISVDGILTTTGYTQEDYTMLGSDDFFYVGGSPSTADL 392

Qy         163 TGGRFSSGITGCVKNLV 179
                |   |:   ||:| :|
Db         393 PGSPVSNNFMGCLKEVV 409

Search completed: March  9, 2004, 17:19:45
Job time : 6.07923 secs
```

```
                    GenCore version 5.1.6
           Copyright (c) 1993 - 2004  Compugen Ltd.

OM protein - protein search, using sw model

Run on:         March  9, 2004, 17:15:08 ; Search time 18.3903 Seconds
                                            (without alignments)
                                            3602.917 Million cell updates/sec

Title:          US-10-006-011A-10
Perfect score:  1115
Sequence:       1 GIAESDWHLEGSGGNDAPGQ..........QPLDLQHRAQAGANTRPCPS 210

Scoring table:  BLOSUM62
                Gapop 10.0 , Gapext 0.5

Searched:       1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters:      1017041

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :       SPTREMBL_25:*
                 1:  sp_archea:*
                 2:  sp_bacteria:*
                 3:  sp_fungi:*
                 4:  sp_human:*
                 5:  sp_invertebrate:*
                 6:  sp_mammal:*
                 7:  sp_mhc:*
                 8:  sp_organelle:*
                 9:  sp_phage:*
                 10:  sp_plant:*
                 11:  sp_rodent:*
                 12:  sp_virus:*
                 13:  sp_vertebrate:*
                 14:  sp_unclassified:*
                 15:  sp_rvirus:*
                 16:  sp_bacteriap:*
                 17:  sp_archeap:*

      Pred. No. is the number of results predicted by chance to have a
      score greater than or equal to the score of the result being printed,
      and is derived by analysis of the total score distribution.

                                 SUMMARIES
                   %
Result            Query
  No.    Score    Match Length DB  ID                      Description
 ----------------------------------------------------------------------------
     1   338.5    30.4     925  5  Q9U7E8                  Q9u7e8 caenorhabdi
     2     292    26.2    1361  5  Q9V7I4                  Q9v7i4 drosophila
     3     292    26.2    1361  5  Q9NGV2                  Q9ngv2 drosophila
     4     289    25.9     478  5  Q86SD6                  Q86sd6 ciona intes
     5   279.5    25.1      68 11   O08591                  O08591 rattus norv
     6   278.5    25.0     152  4  Q8N124                  Q8n124 homo sapien
     7   278.5    25.0     463  4  Q8NAL2                  Q8nal2 homo sapien
     8   278.5    25.0     775  4  Q8N7Y0                  Q8n7y0 homo sapien
     9     276    24.8     152  4  Q8N197                  Q8n197 homo sapien
    10   270.5    24.3     295 11   Q80WX4                  Q80wx4 mus musculu
    11   270.5    24.3    1009 11   Q8BGP3                  Q8bgp3 mus musculu
    12     265    23.8     181  4  Q8N4J5                  Q8n4j5 homo sapien
    13     265    23.8     488  4  Q96IC1                  Q96ic1 homo sapien
    14     265    23.8    2026  4  O00468                  O00468 homo sapien
    15     241    21.6     263 11   Q80V56                  Q80v56 mus musculu
    16     241    21.6     263 11   Q99KT4                  Q99kt4 mus musculu
    17     241    21.6     350 11   Q8K326                  Q8k326 mus musculu
    18     226    20.3     708  5  Q9NFS9                  Q9nfs9 drosophila
    19     226    20.3    1035  5  Q9NEG1                  Q9neg1 drosophila
    20     226    20.3    4117  5  Q8IRV9                  Q8irv9 drosophila
    21     226    20.3    4179  5  Q9W4Y4                  Q9w4y4 drosophila
    22     226    20.3    4223  5  Q8MPN3                  Q8mpn3 drosophila
    23     226    20.3    4228  5  Q8IRV8                  Q8irv8 drosophila
    24   214.5    19.2     822 11   Q8BU56                  Q8bu56 mus musculu
    25   203.5    18.3     313 13   Q9YHF0                  Q9yhf0 rana pipien
    26   200.5    18.0     659 11   Q7TQI9                  Q7tqi9 mus musculu
    27   200.5    18.0     858 11   Q8R145                  Q8r145 mus musculu
    28   185.5    16.6    1518  5  Q21442                  Q21442 caenorhabdi
    29   185.5    16.6    3704  5  P91904                  P91904 caenorhabdi
    30   173.5    15.6     354 11   Q91YG8                  Q91yg8 mus musculu
    31   173.5    15.6    1254 11   Q91VV0                  Q91vv0 mus musculu
    32   170.5    15.3    4307  5  Q19319                  Q19319 caenorhabdi
    33     169    15.2    1560  5  Q18291                  Q18291 caenorhabdi
    34   168.5    15.1    1725  6  Q867A1                  Q867a1 canis famil
    35     168    15.1     119  4  Q9BTD4                  Q9btd4 homo sapien
    36     168    15.1    1392  4  Q8IUE3                  Q8iue3 homo sapien
    37     168    15.1    3102  5  O45614                  O45614 caenorhabdi
    38     166    14.9    1837  5  Q9VCZ9                  Q9vcz9 drosophila
    39   162.5    14.6    1534  4  O75093                  O75093 homo sapien
    40   162.5    14.6    1618  4  Q9UIL7                  Q9uil7 homo sapien
    41   162.5    14.6    1806  4  Q96TG0                  Q96tg0 homo sapien
    42   162.5    14.6    2909  5  Q8MY78                  Q8my78 asterina pe
    43     161    14.4    1100 11   Q8CCT8                  Q8cct8 mus musculu
    44   160.5    14.4     414  4  Q14731                  Q14731 homo sapien
    45   160.5    14.4    3712  5  Q9VRW0                  Q9vrw0 drosophila
```

ALIGNMENTS

```
RESULT 1
Q9U7E8
ID   Q9U7E8      PRELIMINARY;      PRT;   925 AA.
AC   Q9U7E8;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   UNC-52/perlecan (Fragment).
GN   UNC-52.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC   Rhabditidae; Peloderinae; Caenorhabditis.
OX   NCBI_TaxID=6239;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=N2;
RX   MEDLINE=99443780; PubMed=10512861;
RA   Mullen G.P., Rogalski T.M., Bush J.A., Gorji P.R., Moerman D.G.;
RT   "Complex patterns of alternative splicing mediate the spatial and
RT   temporal distribution of Perlecan/UNC-52 in caenorhabditis elegans.";
RL   Mol. Biol. Cell 10:3205-3221(1999).
DR   EMBL; AF132883; AAD25092.1; -.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR006209; EGF_like.
DR   InterPro; IPR006210; IEGF.
DR   InterPro; IPR007110; Ig-like.
DR   InterPro; IPR003598; Ig_c2.
DR   InterPro; IPR001791; Laminin_G.
DR   Pfam; PF00008; EGF; 3.
DR   Pfam; PF00047; ig; 1.
DR   Pfam; PF00054; laminin_G; 2.
DR   SMART; SM00181; EGF; 3.
DR   SMART; SM00408; IGc2; 1.
DR   SMART; SM00282; LamG; 3.
DR   PROSITE; PS00022; EGF_1; 3.
DR   PROSITE; PS50835; IG_LIKE; 1.
DR   PROSITE; PS50025; LAM_G_DOMAIN; 3.
KW   EGF-like domain; Immunoglobulin domain.
FT   NON_TER       1      1
```

```
SQ   SEQUENCE   925 AA;  102174 MW;  474BB5F045D67E0B CRC64;

  Query Match             30.4%;  Score 338.5;  DB 5;  Length 925;
  Best Local Similarity   35.6%;  Pred. No. 5.9e-19;
  Matches   68;  Conservative   38;  Mismatches   76;  Indels    9;  Gaps    2;

Qy          18 PGQYGAYFHDDGFLAFPGHVFSRSLPEVPETIELEVRTSTASGLLLWQGVEVGEAGQGKD 77
               | :: | |  | |:      |     |  | :  : :|   :|:||||| :     | :|
Db         728 PIEHAARFDGDAFIELSSDEFPHLTSEKDEIVAFKFKTEQQNGVLLWQG-QRPTVQQMED 786

Qy          78 FISLGLQDGHLVFRYQLGSGEARLVSEDPINDGEWHRVTALREGRRGSIQVDGEELVSGR 137
               :||:|: :||| | |:|| | | |:||: ::||: | |   |:|| | :::|    | ||
Db         787 YISVGIVNGHLHFSYELGGGAAHLISEERVDDGKEHSVRFERKGREGQMRIDNYREVDGR 846

Qy         138 SPGPNVAVNAKGSVYIGGAPDVATLTGGRFSSGITGCVKNLVLHSARPGAPPPQPLDLQH 197
               | |    :|  |::::|| ||::  ||| ||:   ||: :: |:  :        |||
Db         847 STGILAMLNVDGNIFVGGVPDISKATGGLFSNNFVGCIADVELNGVK--------LDLMA 898

Qy         198 RAQAGANTRPC 208
                |  | | :||
Db         899 TAIDGKNVKPC 909

RESULT 2
Q9V7I4
ID   Q9V7I4      PRELIMINARY;      PRT;  1361 AA.
AC   Q9V7I4;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   CG8403 protein.
GN   SP2353 OR CG8403.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC   Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Berkeley;
RX   MEDLINE=20196006; PubMed=10731132;
RA   Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA   Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA   George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA   Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA   Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA   Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA   Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA   Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA   Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA   Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA   Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA   Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA   de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA   Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA   Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA   Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA   Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA   Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA   Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA   Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA   Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA   Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA   Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA   Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA   Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA   Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA   Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA   Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA   Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA   Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA   Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA   Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA   Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA   Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA   Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA   Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT   "The genome sequence of Drosophila melanogaster.";
RL   Science 287:2185-2195(2000).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Celniker S.E., Adams M.D., Kronmiller B., Wan K.H., Holt R.A.,
RA   Evans C.A., Gocayne J.D., Amanatides P.G., Brandon R.C., Rogers Y.,
RA   Banzon J., An H., Baldwin D., Banzon J., Beeson K.Y., Busam D.A.,
RA   Carlson J.W., Center A., Champe M., Davenport L.B., Dietz S.M.,
RA   Dodson K., Dorsett V., Doup L.E., Doyle C., Dresnek D., Farfan D.,
RA   Ferriera S., Frise E., Galle R.F., Garg N.S., George R.A.,
RA   Gonzalez M., Houck J., Hoskins R.A., Hostin D., Howland T.J.,
RA   Ibegwam C., Jalali M., Kruse D., Li P., Mattei B., Moshrefi A.,
RA   McIntosh T.C., Moy M., Murphy B., Nelson C., Nelson K.A., Nunoo J.,
RA   Pacleb J., Paragas V., Park S., Patel S., Pfeiffer B.,
RA   Phouanenavong S., Pittman G.S., Puri V., Richards S., Scheeler F.,
RA   Stapleton M., Strong R., Svirskas R., Tector C., Tyler D.,
RA   Williams S.M., Zaveri J.S., Smith H.O., Venter J.C., Rubin G.M.;
RT   "Sequencing of Drosophila melanogaster genome.";
RL   Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RA   Misra S., Crosby M.A., Matthews B.B., Bayraktaroglu L., Campbell K.,
RA   Hradecky P., Huang Y., Kaminker J.S., Prochnik S.E., Smith C.D.,
RA   Tupy J.L., Bergman C., Berman B., Carlson J.W., Celniker S.E.,
RA   Clamp M., Drysdale R., Emmert D., Frise E., de Grey A., Harris N.,
RA   Kronmiller B., Marshall B., Millburn G., Richter J., Russo S.,
RA   Searle S.M.J., Smith E., Shu S., Smutniak F., Whitfield E.,
RA   Ashburner M., Gelbart W.M., Rubin G.M., Mungall C.J., Lewis S.E.;
RT   "Annotation of Drosophila melanogaster genome.";
RL   Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
RN   [4]
RP   SEQUENCE FROM N.A.
RA   Adams M.D., Celniker S.E., Gibbs R.A., Rubin G.M., Venter C.J.;
RL   Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
RN   [5]
RP   SEQUENCE FROM N.A.
RA   FlyBase;
RL   Submitted (SEP-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AE003808; AAF58071.2; -.
DR   HSSP; P08709; 1BF9.
DR   FlyBase; FBgn0034070; SP2353.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR006209; EGF_like.
DR   InterPro; IPR006210; IEGF.
DR   InterPro; IPR001791; Laminin_G.
DR   Pfam; PF00008; EGF; 1.
DR   Pfam; PF00054; laminin_G; 3.
DR   SMART; SM00181; EGF; 4.
DR   SMART; SM00282; LamG; 3.
DR   PROSITE; PS00022; EGF_1; 2.
DR   PROSITE; PS01186; EGF_2; 2.
DR   PROSITE; PS50025; LAM_G_DOMAIN; 3.
KW   EGF-like domain.
SQ   SEQUENCE   1361 AA;  150298 MW;  4C71EDA81A12ABED CRC64;

  Query Match             26.2%;  Score 292;  DB 5;  Length 1361;
  Best Local Similarity   35.0%;  Pred. No. 5.9e-15;
  Matches   64;  Conservative   37;  Mismatches   70;  Indels   12;  Gaps    5;

Qy          23 AYFH--DDGFLAFPGHVFSRSLPEVPETIELEVRTSTASGLLLWQGVEVGEAGQGKDFIS 80
               :|||  |   ::    | | |:      : | ::| : :|::|| | : |   :  |::|
Db        1186 SYFHYNDADTMS---QVISYSI-----DLNLRIKTHSENGVILWTGRQ-GTTEEHDDYLS 1236

Qy          81 LGLQDGHLVFRYQLGSGEARL-VSEDPINDGEWHRVTALREGRRGSIQVDGEELVSGRSP 139
               ||:: |:| ||| |||||  :  :   ::|| |||| |:|  : | ::||| : |: |:|
Db        1237 LGIEQGYLHFRYDLGSGEVDIRFNGTKVSDGLWHRVRAIRNSQEGYLEVDGRKTVTLRAP 1296

Qy         140 GPNVAVNAKGSVYIGGAPDVATLTGGRFSSGITGCVKNLVLHSARPGAPPPQPLDLQHRA 199
               |    :|    :|:|| |||   |  |: ||| ||:  :||         |  |  :|
```

```
Db       1297 GKLRQLNTDTGLYVGGMPDVGYFTHQRYFSGIVGCISEIVLAGEMKLNFDPNTLGTEHNV 1356

Qy        200 QAG 202
              : |
Db       1357 ETG 1359

RESULT 3
Q9NGV2
ID   Q9NGV2      PRELIMINARY;      PRT;  1361 AA.
AC   Q9NGV2;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   SP2353.
GN   SP2353 OR CG8403.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC   Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Serano T.L., Pendleton J.D., Rubin G.M.;
RT   "A reverse genetic screen for genes involved in Drosophila
RT   development.";
RL   Submitted (FEB-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF239610; AAF63502.1; -.
DR   HSSP; P08709; 1BF9.
DR   FlyBase; FBgn0034070; SP2353.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR006209; EGF_like.
DR   InterPro; IPR006210; IEGF.
DR   InterPro; IPR001791; Laminin_G.
DR   Pfam; PF00008; EGF; 1.
DR   Pfam; PF00054; laminin_G; 3.
DR   SMART; SM00181; EGF; 3.
DR   SMART; SM00282; LamG; 3.
DR   PROSITE; PS00022; EGF_1; 2.
DR   PROSITE; PS01186; EGF_2; 2.
DR   PROSITE; PS50025; LAM_G_DOMAIN; 3.
KW   EGF-like domain.
SQ   SEQUENCE   1361 AA;  150355 MW;  206D9F9BF9EDFE47 CRC64;

  Query Match             26.2%;  Score 292;  DB 5;  Length 1361;
  Best Local Similarity   35.0%;  Pred. No. 5.9e-15;
  Matches   64;  Conservative   37;  Mismatches   70;  Indels   12;  Gaps    5;

Qy         23 AYFH--DDGFLAFPGHVFSRSLPEVPETIELEVRTSTASGLLLWQGVEVGEAGQGKDFIS 80
              :|||  |   ::    | | |:      : | ::| :|::|| | : |   :  |::|
Db       1186 SYFHYNDADTMS---QVISYSI-----DLNLRIKTHSENGVILWTGRQ-GTTEEHDDYLS 1236

Qy         81 LGLQDGHLVFRYQLGSGEARL-VSEDPINDGEWHRVTALREGRRGSIQVDGEELVSGRSP 139
              ||:: |:| ||| |||||  :  :   ::|| |||| |:|  : | ::||| : |: |:|
Db       1237 LGIEQGYLHFRYDLGSGEVDIRFNGTKVSDGLWHRVRAIRNSQEGYLEVDGRKTVTLRAP 1296

Qy        140 GPNVAVNAKGSVYIGGAPDVATLTGGRFSSGITGCVKNLVLHSARPGAPPPQPLDLQHRA 199
              |    :|    :|:|| |||   |  |: ||| ||:  :||         |  |  :|
Db       1297 GKLRQLNTDTGLYVGGMPDVGYFTHQRYFSGIVGCISEIVLAGEMKLNFDPNTLGTEHNV 1356

Qy        200 QAG 202
              : |
Db       1357 ETG 1359

RESULT 4
Q86SD6
ID   Q86SD6      PRELIMINARY;      PRT;   478 AA.
AC   Q86SD6;
DT   01-JUN-2003 (TrEMBLrel. 24, Created)
DT   01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Perlecan homologue (Fragment).
GN   CI-PERLECAN.
OS   Ciona intestinalis.
OC   Eukaryota; Metazoa; Chordata; Urochordata; Ascidiacea; Enterogona;
OC   Phlebobranchia; Cionidae; Ciona.
OX   NCBI_TaxID=7719;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Hotta K., Takahashi H., Satoh N.;
RT   "Further Characterization of Brachyury-Downstream Genes in Ciona
RT   intestinalis Embryo.";
RL   Submitted (DEC-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AB076890; BAC57513.1; -.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR006209; EGF_like.
DR   InterPro; IPR001791; Laminin_G.
DR   Pfam; PF00054; laminin_G; 2.
DR   SMART; SM00282; LamG; 2.
DR   PROSITE; PS00022; EGF_1; 1.
DR   PROSITE; PS01186; EGF_2; 1.
DR   PROSITE; PS50025; LAM_G_DOMAIN; 2.
FT   NON_TER       1      1
SQ   SEQUENCE   478 AA;  52039 MW;  5D55465CC835B82E CRC64;

  Query Match             25.9%;  Score 289;  DB 5;  Length 478;
  Best Local Similarity   34.2%;  Pred. No. 2.7e-15;
  Matches   65;  Conservative   34;  Mismatches   65;  Indels   26;  Gaps    5;

Qy         17 APGQYGAYFHDDGFLAFPGHVFSRSLPEVPETIELEVRTSTASGLLLWQGVEV------- 69
              :| :: | |  | | :|    |        | |  ::| ::  ||:|: | ::
Db        287 SPPRHLARFSGDSFASFEPTTFPHD--GSAEQISFDMRATSPDGLVLFHGEDLTSEPSET 344

Qy         70 ------GEAGQGKDFISLGLQDGHLVFRYQLGSGEARLVSEDPINDGEWHRVTALREGRR 123
                       |: ||:||:|:|:  ||: : || |: | ||: |: || || || :   ||
Db        345 SRLHKRSHRGRSKDYISVGIQESKLVYSFDLGGGQGRAVSDFPVVDGLWHHVTIV---RR 401

Qy        124 GSIQV----DGEELVSGRSPGPNVAVNAKGSVYIGGAPDVATLTGGRFSSGITGCVKNLV 179
              ||: :    |    |: ::|| :   |  | ||:||||||   | ||:::|  ||: ||
Db        402 GSLALMWLDDVPTPVAAQAPGSHTMANTPGRVYVGGAPDVIFATAGRYTTGFMGCIGNLR 461

Qy        180 LHSARPGAPP 189
                    | ||
Db        462 F----TGIPP 467

RESULT 5
O08591
ID   O08591      PRELIMINARY;      PRT;    68 AA.
AC   O08591;
DT   01-JUL-1997 (TrEMBLrel. 04, Created)
DT   01-JUL-1997 (TrEMBLrel. 04, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Perlecan (Fragment).
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Kidney;
RA   Pyke C., Kristensen P., Ostergaard P.B., Oturai P.S., Romer J.;
RT   "Proteoglycan expression in the normal rat kidney.";
RL   Submitted (OCT-1996) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; U75305; AAB51124.1; -.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR001791; Laminin_G.
DR   Pfam; PF00054; laminin_G; 1.
DR   PROSITE; PS50025; LAM_G_DOMAIN; 1.
FT   NON_TER       1      1
FT   NON_TER      68     68
SQ   SEQUENCE   68 AA;  7395 MW;  5868E45D8A7083E0 CRC64;

  Query Match             25.1%;  Score 279.5;  DB 11;  Length 68;
  Best Local Similarity   79.7%;  Pred. No. 1.3e-15;
```

```
  Matches   55;  Conservative    5;  Mismatches    8;  Indels    1;  Gaps    1;

Qy          18 PGQYGAYFHDDGFLAFPGHVFSRSLPEVPETIELEVRTSTASGLLLWQGVEVGEAGQGKD 77
               |||||||| |:|||| ||: ||||||||||||| |||||||:||||||||   |: : ||
Db           1 PGQYGAYFQDNGFLALPGNSFSRSLPEVPETIEFEVRTSTANGLLLWQGV-AKESSRSKD 59

Qy          78 FISLGLQDG 86
               |||||||||
Db          60 FISLGLQDG 68


RESULT 6
Q8N124
ID   Q8N124      PRELIMINARY;     PRT;   152 AA.
AC   Q8N124;
DT   01-OCT-2002 (TrEMBLrel. 22, Created)
DT   01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Similar to basement membrane-specific heparan sulfate proteoglycan
DE   core protein (HSPG) (Perlecan) (PLC) (Hypothetical protein
DE   FLJ35675).
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Lung;
RA   Strausberg R.;
RL   Submitted (JUN-2002) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Spleen;
RA   Tashiro H., Yamazaki M., Watanabe K., Kumagai A., Itakura S.,
RA   Fukuzumi Y., Fujimori Y., Komiyama M., Sugiyama T., Irie R.,
RA   Otsuki T., Sato H., Ota T., Wakamatsu A., Ishii S., Yamamoto J.,
RA   Isono Y., Kawai-Hio Y., Saito K., Nishikawa T., Kimura K.,
RA   Yamashita H., Matsuo K., Nakamura Y., Sekine M., Kikuchi H., Kanda K.,
RA   Wagatsuma M., Murakawa K., Kanehori K., Takahashi-Fujii A., Oshima A.,
RA   Sugiyama A., Kawakami B., Suzuki Y., Sugano S., Nagahari K.,
RA   Masuho Y., Nagai K., Isogai T.;
RT   "NEDO human cDNA sequencing project.";
RL   Submitted (JUL-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; BC033188; AAH33188.1; -.
DR   EMBL; AK092994; BAC04013.1; -.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR001791; Laminin_G.
DR   Pfam; PF00054; laminin_G; 1.
DR   SMART; SM00282; LamG; 1.
DR   PROSITE; PS50025; LAM_G_DOMAIN; 1.
KW   Hypothetical protein.
SQ   SEQUENCE   152 AA;  16588 MW;  628A79CBD877DB87 CRC64;

  Query Match              25.0%;  Score 278.5;  DB 4;  Length 152;
  Best Local Similarity    37.3%;  Pred. No. 4.3e-15;
  Matches   59;  Conservative   25;  Mismatches   65;  Indels    9;  Gaps    2;

Qy          51 LEVRTSTASGLLLWQGVEVGEAGQGKDFISLGLQDGHLVFRYQLGSGEARLVSEDPINDG 110
               :  :|:   |||||:|          |||||||:|| ||| | |||| | ::     |||
Db           1 MRFKTTAKDGLLLWRG--DSPMRPNSDFISLGLRDGALVFSYNLGSGVASIMVNGSFNDG 58

Qy         111 EWHRVTALREGRRGSIQVDGEELVSGRSPGPNVAVNAKGSVYIGGAPDVATLTGGRFSSG 170
                |||| |:|:|: | | ||     :|:|||    :|  |::|:||  ::|  |   :: |
Db          59 RWHRVKAVRDGQSGKITVDDYGARTGKSPGMMRQLNINGALYVGGMKEIALHTNRQYMRG 118

Qy         171 ITGCVKNLVLHSARPGAPPPQPLDLQHRAQAGANTRPC 208
               : ||: :  | :          : |   |  | |   |
Db         119 LVGCISHFTLST-------DYHISLVEDAVDGKNINTC 149


RESULT 7
Q8NAL2
ID   Q8NAL2      PRELIMINARY;     PRT;   463 AA.
```

```
AC   Q8NAL2;
DT   01-OCT-2002 (TrEMBLrel. 22, Created)
DT   01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Hypothetical protein FLJ35160.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Placenta;
RA   Kawakami B., Sugiyama A., Takemoto M., Sugiyama T., Irie R.,
RA   Otsuki T., Sato H., Wakamatsu A., Ishii S., Yamamoto J., Isono Y.,
RA   Kawai-Hio Y., Saito K., Nishikawa T., Kimura K., Yamashita H.,
RA   Matsuo K., Nakamura Y., Sekine M., Kikuchi H., Kanda K., Wagatsuma M.,
RA   Murakawa K., Kanehori K., Takahashi-Fujii A., Oshima A., Suzuki Y.,
RA   Sugano S., Nagahari K., Masuho Y., Nagai K., Isogai T.;
RT   "NEDO human cDNA sequencing project.";
RL   Submitted (JUL-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AK092479; BAC03900.1; -.
DR   GO; GO:0005509; F:calcium ion binding; IEA.
DR   InterPro; IPR001881; EGF_Ca.
DR   InterPro; IPR006209; EGF_like.
DR   InterPro; IPR006210; IEGF.
DR   InterPro; IPR001791; Laminin_G.
DR   Pfam; PF00008; EGF; 2.
DR   Pfam; PF00054; laminin_G; 2.
DR   SMART; SM00181; EGF; 2.
DR   SMART; SM00179; EGF_CA; 2.
DR   SMART; SM00282; LamG; 2.
DR   PROSITE; PS00022; EGF_1; 2.
DR   PROSITE; PS01186; EGF_2; 2.
DR   PROSITE; PS50025; LAM_G_DOMAIN; 2.
KW   Hypothetical protein; EGF-like domain.
SQ   SEQUENCE   463 AA;  50635 MW;  F688BF2714D5D0EC CRC64;

  Query Match              25.0%;  Score 278.5;  DB 4;  Length 463;
  Best Local Similarity    37.3%;  Pred. No. 1.8e-14;
  Matches   59;  Conservative   25;  Mismatches   65;  Indels    9;  Gaps    2;

Qy          51 LEVRTSTASGLLLWQGVEVGEAGQGKDFISLGLQDGHLVFRYQLGSGEARLVSEDPINDG 110
               :  :|:   |||||:|          |||||||:|| ||| | |||| | ::     |||
Db         312 MRFKTTAKDGLLLWRG--DSPMRPNSDFISLGLRDGALVFSYNLGSGVASIMVNGSFNDG 369

Qy         111 EWHRVTALREGRRGSIQVDGEELVSGRSPGPNVAVNAKGSVYIGGAPDVATLTGGRFSSG 170
                |||| |:|:|: | | ||     :|:|||    :|  |::|:||  ::|  |   :: |
Db         370 RWHRVKAVRDGQSGKITVDDYGARTGKSPGMMRQLNINGALYVGGMKEIALHTNRQYMRG 429

Qy         171 ITGCVKNLVLHSARPGAPPPQPLDLQHRAQAGANTRPC 208
               : ||: :  | :          : |   |  | |   |
Db         430 LVGCISHFTLST-------DYHISLVEDAVDGKNINTC 460


RESULT 8
Q8N7Y0
ID   Q8N7Y0      PRELIMINARY;     PRT;   775 AA.
AC   Q8N7Y0;
DT   01-OCT-2002 (TrEMBLrel. 22, Created)
DT   01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Hypothetical protein FLJ40230.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Testis;
RA   Kawakami B., Sugiyama A., Takemoto M., Sugiyama T., Irie R.,
RA   Otsuki T., Sato H., Wakamatsu A., Ishii S., Yamamoto J., Isono Y.,
RA   Kawai-Hio Y., Saito K., Nishikawa T., Kimura K., Yamashita H.,
RA   Matsuo K., Nakamura Y., Sekine M., Kikuchi H., Kanda K., Wagatsuma M.,
```

```
RA   Murakawa K., Kanehori K., Takahashi-Fujii A., Oshima A., Suzuki Y.,
RA   Sugano S., Nagahari K., Masuho Y., Nagai K., Isogai T.;
RT   "NEDO human cDNA sequencing project.";
RL   Submitted (JUL-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AK097549; BAC05096.1; -.
DR   GO; GO:0005509; F:calcium ion binding; IEA.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR001881; EGF_Ca.
DR   InterPro; IPR006209; EGF_like.
DR   InterPro; IPR006210; IEGF.
DR   InterPro; IPR001791; Laminin_G.
DR   Pfam; PF00008; EGF; 2.
DR   Pfam; PF00054; laminin_G; 3.
DR   SMART; SM00181; EGF; 3.
DR   SMART; SM00179; EGF_CA; 2.
DR   SMART; SM00282; LamG; 3.
DR   PROSITE; PS00022; EGF_1; 3.
DR   PROSITE; PS01186; EGF_2; 2.
DR   PROSITE; PS50025; LAM_G_DOMAIN; 3.
KW   Hypothetical protein; EGF-like domain.
SQ   SEQUENCE   775 AA;  84800 MW;  61A81294F0204ED2 CRC64;

  Query Match             25.0%;  Score 278.5;  DB 4;  Length 775;
  Best Local Similarity   37.3%;  Pred. No. 3.6e-14;
  Matches   59;  Conservative   25;  Mismatches   65;  Indels    9;  Gaps    2;

Qy          51 LEVRTSTASGLLLWQGVEVGEAGQGKDFISLGLQDGHLVFRYQLGSGEARLVSEDPINDG 110
               :  :|:   |||||:|          |||||||:|| ||| | |||| | ::     |||
Db         624 MRFKTTAKDGLLLWRG--DSPMRPNSDFISLGLRDGALVFSYNLGSGVASIMVNGSFNDG 681

Qy         111 EWHRVTALREGRRGSIQVDGEELVSGRSPGPNVAVNAKGSVYIGGAPDVATLTGGRFSSG 170
                |||| |:|:|: | | ||     :|:|||    :|  |::|:||  ::|  |  ::  |
Db         682 RWHRVKAVRDGQSGKITVDDYGARTGKSPGMMRQLNINGALYVGGMKEIALHTNRQYMRG 741

Qy         171 ITGCVKNLVLHSARPGAPPPQPLDLQHRAQAGANTRPC 208
               : ||: :  | :          : |   |  | |   |
Db         742 LVGCISHFTLST-------DYHISLVEDAVDGKNINTC 772


RESULT 9
Q8N197
ID   Q8N197      PRELIMINARY;      PRT;   152 AA.
AC   Q8N197;
DT   01-OCT-2002 (TrEMBLrel. 22, Created)
DT   01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Similar to basement membrane-specific heparan sulfate proteoglycan
DE   core protein (HSPG) (Perlecan) (PLC).
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Kidney;
RA   Strausberg R.;
RL   Submitted (JUN-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; BC033177; AAH33177.1; -.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR001791; Laminin_G.
DR   Pfam; PF00054; laminin_G; 1.
DR   SMART; SM00282; LamG; 1.
DR   PROSITE; PS50025; LAM_G_DOMAIN; 1.
SQ   SEQUENCE   152 AA;  16604 MW;  628A79CBCEADDB87 CRC64;

  Query Match             24.8%;  Score 276;  DB 4;  Length 152;
  Best Local Similarity   41.5%;  Pred. No. 6.9e-15;
  Matches   54;  Conservative   23;  Mismatches   51;  Indels    2;  Gaps    1;

Qy          51 LEVRTSTASGLLLWQGVEVGEAGQGKDFISLGLQDGHLVFRYQLGSGEARLVSEDPINDG 110
               :  :|:   |||||:|          |||||||:|| ||| | |||| | ::     |||
Db           1 MRFKTTAKDGLLLWRG--DSPMRPNSDFISLGLRDGALVFSYNLGSGVASIMVNGSFNDG 58
```

```
Qy         111 EWHRVTALREGRRGSIQVDGEELVSGRSPGPNVAVNAKGSVYIGGAPDVATLTGGRFSSG 170
                |||| |:|:|: | | ||     :|:|||    :|  |::|:||  ::|  |  ::  |
Db          59 RWHRVKAVRDGQSGKITVDDYGARTGKSPGMMRQLNINGALYVGGMKEIALHTNRQYMRG 118

Qy         171 ITGCVKNLVL 180
               : ||: :  |
Db         119 LVGCISHFTL 128


RESULT 10
Q80WX4
ID   Q80WX4      PRELIMINARY;      PRT;   295 AA.
AC   Q80WX4;
DT   01-JUN-2003 (TrEMBLrel. 24, Created)
DT   01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Similar to Agrin (Fragment).
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Olfactory epithelium;
RX   MEDLINE=22388257; PubMed=12477932;
RA   Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA   Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA   Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA   Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA   Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA   Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA   Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA   Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA   Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA   Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA   Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA   Fahey J., Helton E., Ketteman M., Madan A., Rodrigues S., Sanchez A.,
RA   Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA   Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA   Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA   Krzywinski M.I., Skalska U., Smailus D.E., Schnerch A., Schein J.E.,
RA   Jones S.J., Marra M.A.;
RT   "Generation and initial analysis of more than 15,000 full-length human
RT   and mouse cDNA sequences.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Olfactory epithelium;
RA   Strausberg R.;
RL   Submitted (APR-2003) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; BC051455; AAH51455.1; -.
DR   GO; GO:0005509; F:calcium ion binding; IEA.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR000742; EGF_2.
DR   InterPro; IPR001881; EGF_Ca.
DR   InterPro; IPR006209; EGF_like.
DR   InterPro; IPR006210; IEGF.
DR   InterPro; IPR001791; Laminin_G.
DR   Pfam; PF00008; EGF; 1.
DR   Pfam; PF00054; laminin_G; 1.
DR   SMART; SM00181; EGF; 1.
DR   SMART; SM00179; EGF_CA; 1.
DR   SMART; SM00282; LamG; 1.
DR   PROSITE; PS00022; EGF_1; 1.
DR   PROSITE; PS01186; EGF_2; 1.
DR   PROSITE; PS50025; LAM_G_DOMAIN; 2.
FT   NON_TER       1      1
SQ   SEQUENCE   295 AA;  32046 MW;  6165DC666EEC885C CRC64;

  Query Match             24.3%;  Score 270.5;  DB 11;  Length 295;
  Best Local Similarity   36.1%;  Pred. No. 4.6e-14;
  Matches   57;  Conservative   27;  Mismatches   65;  Indels    9;  Gaps    2;
```

```
Qy          51 LEVRTSTASGLLLWQGVEVGEAGQGKDFISLGLQDGHLVFRYQLGSGEARLVSEDPINDG 110
               :  :|:   |||||:|          |||||||:|| |:| | |||| | ::     :||
Db         144 MRFKTTAKDGLLLWRG--DSPMRPNSDFISLGLRDGALIFSYNLGSGVASIMVNGSFSDG 201

Qy         111 EWHRVTALREGRRGSIQVDGEELVSGRSPGPNVAVNAKGSVYIGGAPDVATLTGGRFSSG 170
                |||| |:|:|: | | ||     :|:|||    :|  |::|:||  ::|  |  ::  |
Db         202 RWHRVKAVRDGQSGKITVDDYGARTGKSPGLMRQLNINGALYVGGMKEIALHTNRQYLRG 261

Qy         171 ITGCVKNLVLHSARPGAPPPQPLDLQHRAQAGANTRPC 208
               : ||: :  | :          : |   |  | |   |
Db         262 LVGCISHFTLST-------DYHISLVEDAVDGKNINTC 292


RESULT 11
Q8BGP3
ID   Q8BGP3      PRELIMINARY;      PRT;  1009 AA.
AC   Q8BGP3;
DT   01-MAR-2003 (TrEMBLrel. 23, Created)
DT   01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Hypothetical EGF-like domain.
GN   AU040377.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6J; TISSUE=Forelimb, Skin, Testis, and Thymus;
RX   MEDLINE=22354683; PubMed=12466851;
RA   The FANTOM Consortium,
RA   the RIKEN Genome Exploration Research Group Phase I & II Team;
RT   "Analysis of the mouse transcriptome based on functional annotation of
RT   60,770 full-length cDNAs.";
RL   Nature 420:563-573(2002).
DR   EMBL; AK031136; BAC27271.1; -.
DR   EMBL; AK033332; BAC28235.1; -.
DR   EMBL; AK037223; BAC29762.1; -.
DR   EMBL; AK041546; BAC30982.1; -.
DR   MGD; MGI:2146149; AU040377.
DR   GO; GO:0005509; F:calcium ion binding; IEA.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR000742; EGF_2.
DR   InterPro; IPR001881; EGF_Ca.
DR   InterPro; IPR006209; EGF_like.
DR   InterPro; IPR003962; FnIII_subd.
DR   InterPro; IPR003961; FN_III.
DR   InterPro; IPR008957; FN_III-like.
DR   InterPro; IPR006210; IEGF.
DR   InterPro; IPR001791; Laminin_G.
DR   Pfam; PF00008; EGF; 2.
DR   Pfam; PF00041; fn3; 2.
DR   Pfam; PF00054; laminin_G; 3.
DR   PRINTS; PR00014; FNTYPEIII.
DR   SMART; SM00181; EGF; 3.
DR   SMART; SM00179; EGF_CA; 2.
DR   SMART; SM00060; FN3; 2.
DR   SMART; SM00282; LamG; 3.
DR   PROSITE; PS00022; EGF_1; 3.
DR   PROSITE; PS01186; EGF_2; 2.
DR   PROSITE; PS50025; LAM_G_DOMAIN; 3.
KW   Hypothetical protein.
SQ   SEQUENCE   1009 AA;  109834 MW;  80DD954F255C102B CRC64;

  Query Match             24.3%;  Score 270.5;  DB 11;  Length 1009;
  Best Local Similarity   36.1%;  Pred. No. 2.3e-13;
  Matches   57;  Conservative   27;  Mismatches   65;  Indels    9;  Gaps    2;

Qy          51 LEVRTSTASGLLLWQGVEVGEAGQGKDFISLGLQDGHLVFRYQLGSGEARLVSEDPINDG 110
               :  :|:   |||||:|          |||||||:|| |:| | |||| | ::     :||
Db         858 MRFKTTAKDGLLLWRG--DSPMRPNSDFISLGLRDGALIFSYNLGSGVASIMVNGSFSDG 915

Qy         111 EWHRVTALREGRRGSIQVDGEELVSGRSPGPNVAVNAKGSVYIGGAPDVATLTGGRFSSG 170
```

```
                |||| |:|:|: | | ||     :|:|||    :|  |::|:||  ::|  |  ::  |
Db         916 RWHRVKAVRDGQSGKITVDDYGARTGKSPGLMRQLNINGALYVGGMKEIALHTNRQYLRG 975

Qy         171 ITGCVKNLVLHSARPGAPPPQPLDLQHRAQAGANTRPC 208
               : ||: :  | :          : |   |  | |   |
Db         976 LVGCISHFTLST-------DYHISLVEDAVDGKNINTC 1006


RESULT 12
Q8N4J5
ID   Q8N4J5      PRELIMINARY;      PRT;   181 AA.
AC   Q8N4J5;
DT   01-OCT-2002 (TrEMBLrel. 22, Created)
DT   01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Similar to agrin (Fragment).
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Colon;
RA   Strausberg R.;
RL   Submitted (JUL-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; BC034009; AAH34009.1; -.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR001791; Laminin_G.
DR   Pfam; PF00054; laminin_G; 1.
DR   SMART; SM00282; LamG; 1.
DR   PROSITE; PS50025; LAM_G_DOMAIN; 1.
FT   NON_TER       1      1
SQ   SEQUENCE   181 AA;  19516 MW;  AA91BB3F30EB0F8B CRC64;

  Query Match             23.8%;  Score 265;  DB 4;  Length 181;
  Best Local Similarity   34.2%;  Pred. No. 6.8e-14;
  Matches   65;  Conservative   27;  Mismatches   76;  Indels   22;  Gaps    4;

Qy          31 LAFPGHVFSRSLPEVPET--------IELEVRTSTASGLLLWQGVEVGEAGQGKDFISLG 82
               ||| |  |   |  | |:         || :||    ||:||     |:| :  |:::|
Db           3 LAFDGRTFVEYLNAVTESEKALQSNHFELSLRTEATQGLVLWS----GKATERADYVALA 58

Qy          83 LQDGHLVFRYQLGSGEARLVSEDPINDGEWHRVTALREGRRGSIQVDGEELVSGRSPGPN 142
               : ||||   | |||    | |  |:|   | || | || | ||:||  |  |:| ||
Db          59 IVDGHLQLSYNLGSQPVVLRSTVPVNTNRWLRVVAHREQREGSLQVGNEAPVTGSSPLGA 118

Qy         143 VAVNAKGSVYIGGAPD--VATLTGGRFSSGITGCVKNLVLHSARPGAPPPQPLDLQHRAQ 200
                 ::  |::::|| |:  |       : :|  ||::::|:           || |   |
Db         119 TQLDTDGALWLGGLPELPVGPALPKAYGTGFVGCLRDVVVGR--------HPLHLLEDAV 170

Qy         201 AGANTRPCPS 210
                    ||||:
Db         171 TKPELRPCPT 180


RESULT 13
Q96IC1
ID   Q96IC1      PRELIMINARY;      PRT;   488 AA.
AC   Q96IC1;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Hypothetical protein (Fragment).
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Kidney;
RA   Strausberg R.;
RL   Submitted (MAY-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; BC007649; AAH07649.1; -.
```

```
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR006209; EGF_like.
DR   InterPro; IPR006210; IEGF.
DR   InterPro; IPR001791; Laminin_G.
DR   Pfam; PF00008; EGF; 3.
DR   Pfam; PF00054; laminin_G; 2.
DR   SMART; SM00181; EGF; 2.
DR   SMART; SM00282; LamG; 2.
DR   PROSITE; PS00022; EGF_1; 3.
DR   PROSITE; PS01186; EGF_2; 1.
DR   PROSITE; PS50025; LAM_G_DOMAIN; 2.
KW   Hypothetical protein; EGF-like domain.
FT   NON_TER       1      1
SQ   SEQUENCE   488 AA;  51764 MW;  A732B99CC680E772 CRC64;

  Query Match             23.8%;  Score 265;  DB 4;  Length 488;
  Best Local Similarity   34.2%;  Pred. No. 2.5e-13;
  Matches   65;  Conservative   27;  Mismatches   76;  Indels   22;  Gaps    4;

Qy          31 LAFPGHVFSRSLPEVPET--------IELEVRTSTASGLLLWQGVEVGEAGQGKDFISLG 82
               ||| |  |   |  | |:         || :||    ||:||     |:| :  |:::|
Db         310 LAFDGRTFVEYLNAVTESEKALQSNHFELSLRTEATQGLVLWS----GKATERADYVALA 365

Qy          83 LQDGHLVFRYQLGSGEARLVSEDPINDGEWHRVTALREGRRGSIQVDGEELVSGRSPGPN 142
               : ||||   | |||    | |  |:|   | || | || | ||:||  |  |:| ||
Db         366 IVDGHLQLSYNLGSQPVVLRSTVPVNTNRWLRVVAHREQREGSLQVGNEAPVTGSSPLGA 425

Qy         143 VAVNAKGSVYIGGAPD--VATLTGGRFSSGITGCVKNLVLHSARPGAPPPQPLDLQHRAQ 200
                 ::  |::::|| |:  |       : :|  ||::::|:           || |   |
Db         426 TQLDTDGALWLGGLPELPVGPALPKAYGTGFVGCLRDVVVGR--------HPLHLLEDAV 477

Qy         201 AGANTRPCPS 210
                    ||||:
Db         478 TKPELRPCPT 487


RESULT 14
O00468
ID   O00468      PRELIMINARY;      PRT;  2026 AA.
AC   O00468;
DT   01-JUL-1997 (TrEMBLrel. 04, Created)
DT   01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   AGRIN precursor (Fragment).
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Groffen A.J.A., Buskens C.A.F., Van Kuppevelt T.H.M.S.M.,
RA   Veerkamp J.H., Monnens L.A.H., van den Heuvel L.P.W.J.;
RL   Eur. J. Biochem. 0:0-0(1998).
RN   [2]
RP   SEQUENCE OF 1-153 FROM N.A.
RX   MEDLINE=96224170; PubMed=8617505;
RA   Lennon G., Auffray C., Polymeropoulos M., Soares M.B.;
RT   "The I.M.A.G.E. Consortium: an integrated molecular analysis of
RT   genomes and their expression.";
RL   Genomics 33:151-152(1996).
RN   [3]
RP   SEQUENCE OF 1-153 FROM N.A.
RA   Denzer A.J., Brandenberger R., Gesemann M., Chiquet M., Ruegg M.A.;
RL   J. Cell Biol. 0:0-0(0).
DR   EMBL; AF016903; AAC39776.1; -.
DR   EMBL; U84406; AAB52917.1; -.
DR   HSSP; P00740; 1EDM.
DR   Genew; HGNC:329; AGRN.
DR   GO; GO:0005198; F:structural molecule activity; IEA.
DR   InterPro; IPR004850; Agrin_NtA.
DR   InterPro; IPR006209; EGF_like.
DR   InterPro; IPR003645; FolN.
DR   InterPro; IPR002350; kazal.
DR   InterPro; IPR002049; Laminin_EGF.
DR   InterPro; IPR001791; Laminin_G.
DR   InterPro; IPR000082; SEA_domain.
DR   InterPro; IPR008993; TIMP_like.
DR   InterPro; IPR001455; UPF0033.
DR   Pfam; PF00008; EGF; 4.
DR   Pfam; PF00050; kazal; 9.
DR   Pfam; PF00053; laminin_EGF; 2.
DR   Pfam; PF00054; laminin_G; 3.
DR   Pfam; PF03146; NtA; 1.
DR   Pfam; PF01390; SEA; 1.
DR   PRINTS; PR00011; EGFLAMININ.
DR   SMART; SM00180; EGF_Lam; 2.
DR   SMART; SM00274; FOLN; 5.
DR   SMART; SM00280; KAZAL; 9.
DR   SMART; SM00282; LamG; 3.
DR   SMART; SM00200; SEA; 1.
DR   PROSITE; PS00022; EGF_1; 6.
DR   PROSITE; PS01186; EGF_2; 1.
DR   PROSITE; PS01248; LAMININ_TYPE_EGF; 1.
DR   PROSITE; PS50025; LAM_G_DOMAIN; 3.
DR   PROSITE; PS50024; SEA; 1.
DR   PROSITE; PS01148; UPF0033; 1.
KW   EGF-like domain; Laminin EGF-like domain; Signal.
FT   NON_TER       1      1
FT   SIGNAL       <1     10       POTENTIAL.
FT   CHAIN        11   2026       AGRIN.
SQ   SEQUENCE   2026 AA;  212881 MW;  4AB0EE710CD4B8EF CRC64;

  Query Match             23.8%;  Score 265;  DB 4;  Length 2026;
  Best Local Similarity   34.2%;  Pred. No. 1.6e-12;
  Matches   65;  Conservative   27;  Mismatches   76;  Indels   22;  Gaps    4;

Qy          31 LAFPGHVFSRSLPEVPET--------IELEVRTSTASGLLLWQGVEVGEAGQGKDFISLG 82
               ||| |  |   |  | |:         || :||    ||:||     |:| :  |:::|
Db        1848 LAFDGRTFVEYLNAVTESEKALQSNHFELSLRTEATQGLVLWS----GKATERADYVALA 1903

Qy          83 LQDGHLVFRYQLGSGEARLVSEDPINDGEWHRVTALREGRRGSIQVDGEELVSGRSPGPN 142
               : ||||   | |||    | |  |:|   | || | || | ||:||  |  |:| ||
Db        1904 IVDGHLQLSYNLGSQPVVLRSTVPVNTNRWLRVVAHREQREGSLQVGNEAPVTGSSPLGA 1963

Qy         143 VAVNAKGSVYIGGAPD--VATLTGGRFSSGITGCVKNLVLHSARPGAPPPQPLDLQHRAQ 200
                 ::  |::::|| |:  |       : :|  ||::::|:           || |   |
Db        1964 TQLDTDGALWLGGLPELPVGPALPKAYGTGFVGCLRDVVVGR--------HPLHLLEDAV 2015

Qy         201 AGANTRPCPS 210
                    ||||:
Db        2016 TKPELRPCPT 2025


RESULT 15
Q80V56
ID   Q80V56      PRELIMINARY;      PRT;   263 AA.
AC   Q80V56;
DT   01-JUN-2003 (TrEMBLrel. 24, Created)
DT   01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Agrn protein (Fragment).
GN   AGRN.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=FVB/N; TISSUE=Breast tumor;
RX   MEDLINE=22388257; PubMed=12477932;
RA   Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA   Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA   Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA   Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA   Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA   Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
```

```
RA   Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA   Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA   Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA   Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA   Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA   Fahey J., Helton E., Ketteman M., Madan A., Rodrigues S., Sanchez A.,
RA   Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA   Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA   Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA   Krzywinski M.I., Skalska U., Smailus D.E., Schnerch A., Schein J.E.,
RA   Jones S.J., Marra M.A.;
RT   "Generation and initial analysis of more than 15,000 full-length human
RT   and mouse cDNA sequences.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=FVB/N; TISSUE=Breast tumor;
RA   Strausberg R.;
RL   Submitted (JAN-2003) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; BC043318; AAH43318.1; -.
DR   GO; GO:0005509; F:calcium ion binding; IEA.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR001881; EGF_Ca.
DR   InterPro; IPR006209; EGF_like.
DR   InterPro; IPR006210; IEGF.
DR   InterPro; IPR001791; Laminin_G.
DR   Pfam; PF00008; EGF; 1.
DR   Pfam; PF00054; laminin_G; 1.
DR   SMART; SM00181; EGF; 1.
DR   SMART; SM00179; EGF_CA; 1.
DR   SMART; SM00282; LamG; 1.
DR   PROSITE; PS00022; EGF_1; 1.
DR   PROSITE; PS01186; EGF_2; 1.
DR   PROSITE; PS50025; LAM_G_DOMAIN; 1.
FT   NON_TER       1      1
SQ   SEQUENCE   263 AA;  28395 MW;  89708020EBFBD515 CRC64;

  Query Match            21.6%;  Score 241;  DB 11;  Length 263;
  Best Local Similarity  30.0%;  Pred. No. 1e-11;
  Matches   64;  Conservative   31;  Mismatches   86;  Indels   32;  Gaps    5;

Qy         18 PGQYGAYFHDDGF----------LAFPGHVFSRSLPEVPET--------IELEVRTSTAS 59
              || :     : |           ||| |  :   |  | |:         || :||
Db         62 PGGFSGLHCEKGIVEKSVGDLETLAFDGRTYIEYLNAVTESEKALQSNHFELSLRTEATQ 121

Qy         60 GLLLWQGVEVGEAGQGKDFISLGLQDGHLVFRYQLGSGEARLVSEDPINDGEWHRVTALR 119
              ||:||    :|: |:  |:::| : ||||   | |||    | |   :|   | || | |
Db        122 GLVLW----IGKVGERADYMALAIVDGHLQLSYDLGSQPVVLRSTVKVNTNRWLRVRAHR 177

Qy        120 EGRRGSIQVDGEELVSGRSPGPNVAVNAKGSVYIGGAP--DVATLTGGRFSSGITGCVKN 177
              | | ||:||  | |:| ||     ::  |::::||     |       : :|  ||:::
Db        178 EHREGSLQVGNEAPVTGSSPLGATQLDTDGALWLGGLQKLPVGQALPKAYGTGFVGCLRD 237

Qy        178 LVLHSARPGAPPPQPLDLQHRAQAGANTRPCPS 210
              :|:          : | |   |      ||||:
Db        238 VVVGH--------RQLHLLEDAVTKPELRPCPT 262

Search completed: March  9, 2004, 17:21:44
Job time : 19.3903 secs
```

```
                         GenCore version 5.1.6
                 Copyright (c) 1993 - 2004  Compugen Ltd.

OM protein - protein search, using sw model

Run on:         March  9, 2004, 17:19:48 ; Search time 25.5713 Seconds
                                           (without alignments)
                                           2320.373 Million cell updates/sec

Title:          US-10-006-011A-10
Perfect score:  210
Sequence:       1 GIAESDWHLEGSGGNDAPGQ..........QPLDLQHRAQAGANTRPCPS 210

Scoring table:  OLIGO
                Gapop 60.0 , Gapext 60.0

Searched:       1586107 seqs, 282547505 residues

Word size :       0

Total number of hits satisfying chosen parameters:      1586107

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :       A_Geneseq_29Jan04:*
                1:  geneseqp1980s:*
                2:  geneseqp1990s:*
                3:  geneseqp2000s:*
                4:  geneseqp2001s:*
                5:  geneseqp2002s:*
                6:  geneseqp2003as:*
                7:  geneseqp2003bs:*
                8:  geneseqp2004s:*

      Pred. No. is the number of results predicted by chance to have a
      score greater than or equal to the score of the result being printed,
      and is derived by analysis of the total score distribution.

                               SUMMARIES
                     %
Result             Query
  No.     Score    Match Length DB  ID                        Description
 ---------------------------------------------------------------------------
     1      210    100.0   4391  6  AAE34390                  Aae34390 Human per
     2      195     92.9    195  4  AAB31890                  Aab31890 Amino aci
     3      150     71.4   4393  4  AAB31889                  Aab31889 Amino aci
     4      102     48.6   4436  4  ABG23265                  Abg23265 Novel hum
     5       18      8.6     18  4  ABB55911                  Abb55911 Vascular
     6       14      6.7     14  4  ABB55908                  Abb55908 Vascular
     7       14      6.7     14  4  ABB55905                  Abb55905 Vascular
     8       10      4.8     10  4  ABB55910                  Abb55910 Vascular
     9        9      4.3     15  4  ABB55909                  Abb55909 Vascular
    10        9      4.3    159  6  ABM65732                  Abm65732 Propionib
    11        9      4.3   3319  4  ABB70376                  Abb70376 Drosophil
    12        8      3.8      8  4  ABB55907                  Abb55907 Vascular
    13        8      3.8     14  4  ABB55906                  Abb55906 Vascular
    14        8      3.8     54  4  AAU40544                  Aau40544 Propionib
    15        8      3.8     54  6  ABM37063                  Abm37063 Propionib
    16        8      3.8     80  4  AAM16081                  Aam16081 Peptide #
    17        8      3.8     80  4  ABB35070                  Abb35070 Peptide #
    18        8      3.8     80  4  AAM28574                  Aam28574 Peptide #
    19        8      3.8     80  4  ABB29890                  Abb29890 Peptide #
    20        8      3.8     80  4  ABB20486                  Abb20486 Protein #
    21        8      3.8     80  4  AAM68258                  Aam68258 Human bon
    22        8      3.8     80  4  AAM55885                  Aam55885 Human bra
    23        8      3.8     80  4  ABG49912                  Abg49912 Human liv
    24        8      3.8     80  4  AAM03808                  Aam03808 Peptide #
    25        8      3.8     80  5  ABG37793                  Abg37793 Human pep
    26        8      3.8    161  4  ABG20412                  Abg20412 Novel hum
    27        8      3.8    445  4  AAU42626                  Aau42626 Propionib
    28        8      3.8    445  6  ABM39145                  Abm39145 Propionib
    29        8      3.8    494  6  ABU18011                  Abu18011 Protein e
    30        8      3.8   1130  2  AAR13436                  Aar13436 Merosin M
    31        8      3.8   1130  2  AAR71729                  Aar71729 Merosin m
    32        8      3.8   3070  5  AAO17359                  Aao17359 Human lam
    33        8      3.8   3084  3  AAB19796                  Aab19796 Mouse lam
    34        8      3.8   3088  3  AAB19794                  Aab19794 Human lam
    35        8      3.8   3089  3  AAB19792                  Aab19792 Human lam
    36        8      3.8   3106  3  AAB19795                  Aab19795 Mouse lam
    37        8      3.8   3106  7  ADE61792                  Ade61792 Rat Prote
    38        8      3.8   3110  2  AAR71730                  Aar71730 Merosin m
    39        8      3.8   3110  2  AAY15460                  Aay15460 Human lam
    40        8      3.8   3110  3  AAB19793                  Aab19793 Human lam
    41        8      3.8   3110  3  AAB19791                  Aab19791 Human lam
    42        8      3.8   3110  5  AAU84345                  Aau84345 Protein L
    43        8      3.8   3122  7  ADE61794                  Ade61794 Human Pro
    44        8      3.8   3150  4  ABG20414                  Abg20414 Novel hum
    45        7      3.3      7  4  AAM44973                  Aam44973 H11 bindi

                                  ALIGNMENTS

RESULT 1
AAE34390
ID   AAE34390 standard; protein; 4391 AA.
XX
AC   AAE34390;
XX
DT   14-MAY-2003  (first entry)
XX
DE   Human perlecan protein.
XX
KW   Human; diagnosis; osteoarthritis; rheumatoid arthritis; perlecan.
XX
OS   Homo sapiens.
XX
PN   WO200295415-A2.
XX
PD   28-NOV-2002.
XX
PF   22-MAY-2002; 2002WO-EP005612.
XX
PR   23-MAY-2001; 2001GB-00012626.
XX
PA   (OSTE-) OSTEOMETER BIO TECH AS.
XX
PI   Christgau S,  Henriksen DB,  Cloos PAC;
XX
DR   WPI; 2003-140389/13.
XX
PT   An assay for the diagnosis or assessment of the severity of
PT   osteoarthritis or rheumatoid arthritis comprising detecting an isomerized
PT   or optically inverted protein in a sample.
XX
PS   Disclosure; Page 46-67; 106pp; English.
XX
CC   The invention relates to an assay for the diagnosis or assessment of the
CC   severity of osteoarthritis or rheumatoid arthritis. The assay involves
CC   measuring (in a biological sample) the amount or presence of an
CC   isomerised or optically inverted protein or one or more isomerised or
CC   optically inverted fragments from proteins such as perlecan, biglycan,
CC   decorin, fibrillin-1 or protocadherin. The assay is useful for the
CC   diagnosis or assessment of the severity of osteoarthritis or rheumatoid
CC   arthritis. The present sequence is human perlecan protein
XX
SQ   Sequence 4391 AA;

  Query Match             100.0%;  Score 210;  DB 6;  Length 4391;
  Best Local Similarity   100.0%;  Pred. No. 2e-200;
  Matches  210;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;
```

```
Qy          1 GIAESDWHLEGSGGNDAPGQYGAYFHDDGFLAFPGHVFSRSLPEVPETIELEVRTSTASG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4182 GIAESDWHLEGSGGNDAPGQYGAYFHDDGFLAFPGHVFSRSLPEVPETIELEVRTSTASG 4241

Qy         61 LLLWQGVEVGEAGQGKDFISLGLQDGHLVFRYQLGSGEARLVSEDPINDGEWHRVTALRE 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4242 LLLWQGVEVGEAGQGKDFISLGLQDGHLVFRYQLGSGEARLVSEDPINDGEWHRVTALRE 4301

Qy        121 GRRGSIQVDGEELVSGRSPGPNVAVNAKGSVYIGGAPDVATLTGGRFSSGITGCVKNLVL 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4302 GRRGSIQVDGEELVSGRSPGPNVAVNAKGSVYIGGAPDVATLTGGRFSSGITGCVKNLVL 4361

Qy        181 HSARPGAPPPQPLDLQHRAQAGANTRPCPS 210
              ||||||||||||||||||||||||||||||
Db       4362 HSARPGAPPPQPLDLQHRAQAGANTRPCPS 4391


RESULT 2
AAB31890
ID   AAB31890 standard; protein; 195 AA.
XX
AC   AAB31890;
XX
DT   15-MAY-2001  (first entry)
XX
DE   Amino acid sequence of the C-terminal of the human perlecan protein.
XX
KW   Human; perlecan; retinol-binding plasma protein; calgranulin B; vaccine;
KW   ganglioside GM2 activator; saposin B; degenerative disease; glial cell;
KW   neurological disease; auto-immune disease; multiple sclerosis; toxicity;
KW   Alzheimer's disease; Parkinson's disease; amyotrophic lateral sclerosis;
KW   rheumatoid polyarthritis; lupus erythematosus; gene therapy.
XX
OS   Homo sapiens.
XX
PN   WO200105422-A2.
XX
PD   25-JAN-2001.
XX
PF   17-JUL-2000; 2000WO-FR002057.
XX
PR   15-JUL-1999;   99FR-00009372.
XX
PA   (INMR ) BIOMERIEUX STELHYS.
XX
PI   Roecklin D,  Kolbe H,  Charles M,  Malcus C,  Santoro L,  Perron H;
XX
DR   WPI; 2001-159475/16.
DR   N-PSDB; AAF54728.
XX
PT   Detecting, preventing and treating degenerative, neurological and
PT   autoimmune diseases, particularly multiple sclerosis, using specified
PT   polypeptides or related nucleic acid or ligand.
XX
PS   Claim 1; Page 152-153; 209pp; French.
XX
CC   The present sequence represents a human polypeptide, which is used in the
CC   method of the invention. The specification describes a method which uses
CC   at least one polypeptide or polynucleotide sequence belonging to the
CC   perlecan, precursor of the retinol-binding plasma protein, precursor of
CC   the ganglioside GM2 activator, calgranulin B or saposin B protein
CC   families. The method is used for detecting, preventing or treating a
CC   degenerative, neurological and/or auto-immune disease. The
CC   polynucleotides and polypeptides are used for diagnosis, prognosis,
CC   prevention and treatment of multiple sclerosis (in its various forms and
CC   phases). They may also be useful in cases of e.g. Alzheimer's and
CC   Parkinson's diseases, amyotrophic lateral sclerosis, rheumatoid
CC   polyarthritis and lupus erythematosus, including use as vaccines and in
CC   gene therapy (expression of sense or antisense sequences). They can also
CC   be used to assess efficacy of potential therapeutic agents, particularly
CC   compounds that reduce or inhibit toxicity towards glial cells
XX
```

```
SQ   Sequence 195 AA;

  Query Match             92.9%;  Score 195;  DB 4;  Length 195;
  Best Local Similarity   100.0%;  Pred. No. 1.3e-186;
  Matches  195;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         16 DAPGQYGAYFHDDGFLAFPGHVFSRSLPEVPETIELEVRTSTASGLLLWQGVEVGEAGQG 75
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 DAPGQYGAYFHDDGFLAFPGHVFSRSLPEVPETIELEVRTSTASGLLLWQGVEVGEAGQG 60

Qy         76 KDFISLGLQDGHLVFRYQLGSGEARLVSEDPINDGEWHRVTALREGRRGSIQVDGEELVS 135
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 KDFISLGLQDGHLVFRYQLGSGEARLVSEDPINDGEWHRVTALREGRRGSIQVDGEELVS 120

Qy        136 GRSPGPNVAVNAKGSVYIGGAPDVATLTGGRFSSGITGCVKNLVLHSARPGAPPPQPLDL 195
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GRSPGPNVAVNAKGSVYIGGAPDVATLTGGRFSSGITGCVKNLVLHSARPGAPPPQPLDL 180

Qy        196 QHRAQAGANTRPCPS 210
              |||||||||||||||
Db        181 QHRAQAGANTRPCPS 195


RESULT 3
AAB31889
ID   AAB31889 standard; protein; 4393 AA.
XX
AC   AAB31889;
XX
DT   15-MAY-2001  (first entry)
XX
DE   Amino acid sequence of a human protein.
XX
KW   Human; perlecan; retinol-binding plasma protein; calgranulin B; vaccine;
KW   ganglioside GM2 activator; saposin B; degenerative disease; glial cell;
KW   neurological disease; auto-immune disease; multiple sclerosis; toxicity;
KW   Alzheimer's disease; Parkinson's disease; amyotrophic lateral sclerosis;
KW   rheumatoid polyarthritis; lupus erythematosus; gene therapy.
XX
OS   Homo sapiens.
XX
PN   WO200105422-A2.
XX
PD   25-JAN-2001.
XX
PF   17-JUL-2000; 2000WO-FR002057.
XX
PR   15-JUL-1999;   99FR-00009372.
XX
PA   (INMR ) BIOMERIEUX STELHYS.
XX
PI   Roecklin D,  Kolbe H,  Charles M,  Malcus C,  Santoro L,  Perron H;
XX
DR   WPI; 2001-159475/16.
XX
PT   Detecting, preventing and treating degenerative, neurological and
PT   autoimmune diseases, particularly multiple sclerosis, using specified
PT   polypeptides or related nucleic acid or ligand.
XX
PS   Claim 1; Page 138-152; 209pp; French.
XX
CC   The present sequence represents a human protein, which is used in the
CC   method of the invention. The specification describes a method which uses
CC   at least one polypeptide or polynucleotide sequence belonging to the
CC   perlecan, precursor of the retinol-binding plasma protein, precursor of
CC   the ganglioside GM2 activator, calgranulin B or saposin B protein
CC   families. The method is used for detecting, preventing or treating a
CC   degenerative, neurological and/or auto-immune disease. The
CC   polynucleotides and polypeptides are used for diagnosis, prognosis,
CC   prevention and treatment of multiple sclerosis (in its various forms and
CC   phases). They may also be useful in cases of e.g. Alzheimer's and
CC   Parkinson's diseases, amyotrophic lateral sclerosis, rheumatoid
CC   polyarthritis and lupus erythematosus, including use as vaccines and in
```

```
CC   gene therapy (expression of sense or antisense sequences). They can also
CC   be used to assess efficacy of potential therapeutic agents, particularly
CC   compounds that reduce or inhibit toxicity towards glial cells
XX
SQ   Sequence 4393 AA;

  Query Match             71.4%;  Score 150;  DB 4;  Length 4393;
  Best Local Similarity   100.0%;  Pred. No. 2.3e-140;
  Matches  150;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 GIAESDWHLEGSGGNDAPGQYGAYFHDDGFLAFPGHVFSRSLPEVPETIELEVRTSTASG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4184 GIAESDWHLEGSGGNDAPGQYGAYFHDDGFLAFPGHVFSRSLPEVPETIELEVRTSTASG 4243

Qy         61 LLLWQGVEVGEAGQGKDFISLGLQDGHLVFRYQLGSGEARLVSEDPINDGEWHRVTALRE 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4244 LLLWQGVEVGEAGQGKDFISLGLQDGHLVFRYQLGSGEARLVSEDPINDGEWHRVTALRE 4303

Qy        121 GRRGSIQVDGEELVSGRSPGPNVAVNAKGS 150
              ||||||||||||||||||||||||||||||
Db       4304 GRRGSIQVDGEELVSGRSPGPNVAVNAKGS 4333


RESULT 4
ABG23265
ID   ABG23265 standard; protein; 4436 AA.
XX
AC   ABG23265;
XX
DT   18-FEB-2002  (first entry)
XX
DE   Novel human diagnostic protein #23256.
XX
KW   Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW   food supplement; medical imaging; diagnostic; genetic disorder.
XX
OS   Homo sapiens.
XX
PN   WO200175067-A2.
XX
PD   11-OCT-2001.
XX
PF   30-MAR-2001; 2001WO-US008631.
XX
PR   31-MAR-2000; 2000US-00540217.
PR   23-AUG-2000; 2000US-00649167.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Drmanac RT,  Liu C,  Tang YT;
XX
DR   WPI; 2001-639362/73.
DR   N-PSDB; AAS87452.
XX
PT   New isolated polynucleotide and encoded polypeptides, useful in
PT   diagnostics, forensics, gene mapping, identification of mutations
PT   responsible for genetic disorders or other traits and to assess
PT   biodiversity.
XX
PS   Claim 20; SEQ ID NO 53624; 103pp; English.
XX
CC   The invention relates to isolated polynucleotide (I) and polypeptide (II)
CC   sequences. (I) is useful as hybridisation probes, polymerase chain
CC   reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
CC   and in recombinant production of (II). The polynucleotides are also used
CC   in diagnostics as expressed sequence tags for identifying expressed
CC   genes. (I) is useful in gene therapy techniques to restore normal
CC   activity of (II) or to treat disease states involving (II). (II) is
CC   useful for generating antibodies against it, detecting or quantitating a
CC   polypeptide in tissue, as molecular weight markers and as a food
CC   supplement. (II) and its binding partners are useful in medical imaging
CC   of sites expressing (II). (I) and (II) are useful for treating disorders
CC   involving aberrant protein expression or biological activity. The
```

```
CC   polypeptide and polynucleotide sequences have applications in
CC   diagnostics, forensics, gene mapping, identification of mutations
CC   responsible for genetic disorders or other traits to assess biodiversity
CC   and to produce other types of data and products dependent on DNA and
CC   amino acid sequences. ABG00010-ABG30377 represent novel human diagnostic
CC   amino acid sequences of the invention. Note: The sequence data for this
CC   patent did not appear in the printed specification, but was obtained in
CC   electronic format directly from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 4436 AA;

  Query Match             48.6%;  Score 102;  DB 4;  Length 4436;
  Best Local Similarity   100.0%;  Pred. No. 2.6e-92;
  Matches  102;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        108 NDGEWHRVTALREGRRGSIQVDGEELVSGRSPGPNVAVNAKGSVYIGGAPDVATLTGGRF 167
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4334 NDGEWHRVTALREGRRGSIQVDGEELVSGRSPGPNVAVNAKGSVYIGGAPDVATLTGGRF 4393

Qy        168 SSGITGCVKNLVLHSARPGAPPPQPLDLQHRAQAGANTRPCP 209
              ||||||||||||||||||||||||||||||||||||||||||
Db       4394 SSGITGCVKNLVLHSARPGAPPPQPLDLQHRAQAGANTRPCP 4435


RESULT 5
ABB55911
ID   ABB55911 standard; peptide; 18 AA.
XX
AC   ABB55911;
XX
DT   15-FEB-2002  (first entry)
XX
DE   Vascular dementia-associated protein isoform (VPI) 111.
XX
KW   Vascular Dementia; VD; VD-associated protein isoform; VPI; screening;
KW   diagnosis; prognosis; gene therapy.
XX
OS   Homo sapiens.
XX
PN   WO200169261-A2.
XX
PD   20-SEP-2001.
XX
PF   14-MAR-2001; 2001WO-GB001106.
XX
PR   15-MAR-2000; 2000GB-00006285.
PR   24-NOV-2000; 2000GB-00028734.
PR   28-NOV-2000; 2000US-00724391.
XX
PA   (OXFO-) OXFORD GLYCOSCIENCES UK LTD.
XX
PI   Herath HMAC,  Parekh RB,  Rohlff C;
XX
DR   WPI; 2001-557937/62.
XX
PT   Screening, diagnosis or prognosis of vascular dementia (VD), useful for
PT   determining stage of VD and monitoring the effect of VD therapy,
PT   comprises analyzing body fluid by 2-dimensional electrophoresis for
PT   features correlated with VD.
XX
PS   Claim 6; Page 32; 151pp; English.
XX
CC   The invention relates to screening, diagnosis or prognosis of Vascular
CC   Dementia (VD) in a subject comprising analysing body fluid from the
CC   subject by 2-dimensional (2-D) electrophoresis to generate a 2-D array of
CC   features containing at least one chosen feature whose relative abundance
CC   correlates with the presence, absence, stage or severity of VD or
CC   predicts the onset or course of VD, especially detecting in a sample of
CC   cerebrospinal fluid (CSF) from the subject one of 223 VD-associated
CC   protein isoforms (VPIs) (ABB55801-ABB56295) as fully defined in the
CC   specification. Detecting VD-associated features and VPI is useful for the
CC   screening, diagnosis or prognosis of VD, for determining the stage or
```

```
CC   severity of VD, for identifying a subject at risk of VD or for monitoring
CC   the effect of therapy administered to a subject having VD. Nucleic acids
CC   encoding a VPI or inhibiting the function of a VPI are useful for the
CC   treatment of VD and for gene therapy
XX
SQ   Sequence 18 AA;

  Query Match             8.6%;  Score 18;  DB 4;  Length 18;
  Best Local Similarity   100.0%;  Pred. No. 2.3e-10;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        149 GSVYIGGAPDVATLTGGR 166
              ||||||||||||||||||
Db          1 GSVYIGGAPDVATLTGGR 18


RESULT 6
ABB55908
ID   ABB55908 standard; peptide; 14 AA.
XX
AC   ABB55908;
XX
DT   15-FEB-2002  (first entry)
XX
DE   Vascular dementia-associated protein isoform (VPI) 108.
XX
KW   Vascular Dementia; VD; VD-associated protein isoform; VPI; screening;
KW   diagnosis; prognosis; gene therapy.
XX
OS   Homo sapiens.
XX
PN   WO200169261-A2.
XX
PD   20-SEP-2001.
XX
PF   14-MAR-2001; 2001WO-GB001106.
XX
PR   15-MAR-2000; 2000GB-00006285.
PR   24-NOV-2000; 2000GB-00028734.
PR   28-NOV-2000; 2000US-00724391.
XX
PA   (OXFO-) OXFORD GLYCOSCIENCES UK LTD.
XX
PI   Herath HMAC,  Parekh RB,  Rohlff C;
XX
DR   WPI; 2001-557937/62.
XX
PT   Screening, diagnosis or prognosis of vascular dementia (VD), useful for
PT   determining stage of VD and monitoring the effect of VD therapy,
PT   comprises analyzing body fluid by 2-dimensional electrophoresis for
PT   features correlated with VD.
XX
PS   Claim 6; Page 32; 151pp; English.
XX
CC   The invention relates to screening, diagnosis or prognosis of Vascular
CC   Dementia (VD) in a subject comprising analysing body fluid from the
CC   subject by 2-dimensional (2-D) electrophoresis to generate a 2-D array of
CC   features containing at least one chosen feature whose relative abundance
CC   correlates with the presence, absence, stage or severity of VD or
CC   predicts the onset or course of VD, especially detecting in a sample of
CC   cerebrospinal fluid (CSF) from the subject one of 223 VD-associated
CC   protein isoforms (VPIs) (ABB55801-ABB56295) as fully defined in the
CC   specification. Detecting VD-associated features and VPI is useful for the
CC   screening, diagnosis or prognosis of VD, for determining the stage or
CC   severity of VD, for identifying a subject at risk of VD or for monitoring
CC   the effect of therapy administered to a subject having VD. Nucleic acids
CC   encoding a VPI or inhibiting the function of a VPI are useful for the
CC   treatment of VD and for gene therapy
XX
SQ   Sequence 14 AA;

  Query Match             6.7%;  Score 14;  DB 4;  Length 14;
  Best Local Similarity   100.0%;  Pred. No. 1.9e-06;
```

```
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        101 LVSEDPINDGEWHR 114
              ||||||||||||||
Db          1 LVSEDPINDGEWHR 14


RESULT 7
ABB55905
ID   ABB55905 standard; peptide; 14 AA.
XX
AC   ABB55905;
XX
DT   15-FEB-2002  (first entry)
XX
DE   Vascular dementia-associated protein isoform (VPI) 105.
XX
KW   Vascular Dementia; VD; VD-associated protein isoform; VPI; screening;
KW   diagnosis; prognosis; gene therapy.
XX
OS   Homo sapiens.
XX
PN   WO200169261-A2.
XX
PD   20-SEP-2001.
XX
PF   14-MAR-2001; 2001WO-GB001106.
XX
PR   15-MAR-2000; 2000GB-00006285.
PR   24-NOV-2000; 2000GB-00028734.
PR   28-NOV-2000; 2000US-00724391.
XX
PA   (OXFO-) OXFORD GLYCOSCIENCES UK LTD.
XX
PI   Herath HMAC,  Parekh RB,  Rohlff C;
XX
DR   WPI; 2001-557937/62.
XX
PT   Screening, diagnosis or prognosis of vascular dementia (VD), useful for
PT   determining stage of VD and monitoring the effect of VD therapy,
PT   comprises analyzing body fluid by 2-dimensional electrophoresis for
PT   features correlated with VD.
XX
PS   Claim 6; Page 32; 151pp; English.
XX
CC   The invention relates to screening, diagnosis or prognosis of Vascular
CC   Dementia (VD) in a subject comprising analysing body fluid from the
CC   subject by 2-dimensional (2-D) electrophoresis to generate a 2-D array of
CC   features containing at least one chosen feature whose relative abundance
CC   correlates with the presence, absence, stage or severity of VD or
CC   predicts the onset or course of VD, especially detecting in a sample of
CC   cerebrospinal fluid (CSF) from the subject one of 223 VD-associated
CC   protein isoforms (VPIs) (ABB55801-ABB56295) as fully defined in the
CC   specification. Detecting VD-associated features and VPI is useful for the
CC   screening, diagnosis or prognosis of VD, for determining the stage or
CC   severity of VD, for identifying a subject at risk of VD or for monitoring
CC   the effect of therapy administered to a subject having VD. Nucleic acids
CC   encoding a VPI or inhibiting the function of a VPI are useful for the
CC   treatment of VD and for gene therapy
XX
SQ   Sequence 14 AA;

  Query Match             6.7%;  Score 14;  DB 4;  Length 14;
  Best Local Similarity   100.0%;  Pred. No. 1.9e-06;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        124 GSIQVDGEELVSGR 137
              ||||||||||||||
Db          1 GSIQVDGEELVSGR 14


RESULT 8
ABB55910
```

```
ID   ABB55910 standard; peptide; 10 AA.
XX
AC   ABB55910;
XX
DT   15-FEB-2002  (first entry)
XX
DE   Vascular dementia-associated protein isoform (VPI) 110.
XX
KW   Vascular Dementia; VD; VD-associated protein isoform; VPI; screening;
KW   diagnosis; prognosis; gene therapy.
XX
OS   Homo sapiens.
XX
PN   WO200169261-A2.
XX
PD   20-SEP-2001.
XX
PF   14-MAR-2001; 2001WO-GB001106.
XX
PR   15-MAR-2000; 2000GB-00006285.
PR   24-NOV-2000; 2000GB-00028734.
PR   28-NOV-2000; 2000US-00724391.
XX
PA   (OXFO-) OXFORD GLYCOSCIENCES UK LTD.
XX
PI   Herath HMAC,  Parekh RB,  Rohlff C;
XX
DR   WPI; 2001-557937/62.
XX
PT   Screening, diagnosis or prognosis of vascular dementia (VD), useful for
PT   determining stage of VD and monitoring the effect of VD therapy,
PT   comprises analyzing body fluid by 2-dimensional electrophoresis for
PT   features correlated with VD.
XX
PS   Claim 6; Page 32; 151pp; English.
XX
CC   The invention relates to screening, diagnosis or prognosis of Vascular
CC   Dementia (VD) in a subject comprising analysing body fluid from the
CC   subject by 2-dimensional (2-D) electrophoresis to generate a 2-D array of
CC   features containing at least one chosen feature whose relative abundance
CC   correlates with the presence, absence, stage or severity of VD or
CC   predicts the onset or course of VD, especially detecting in a sample of
CC   cerebrospinal fluid (CSF) from the subject one of 223 VD-associated
CC   protein isoforms (VPIs) (ABB55801-ABB56295) as fully defined in the
CC   specification. Detecting VD-associated features and VPI is useful for the
CC   screening, diagnosis or prognosis of VD, for determining the stage or
CC   severity of VD, for identifying a subject at risk of VD or for monitoring
CC   the effect of therapy administered to a subject having VD. Nucleic acids
CC   encoding a VPI or inhibiting the function of a VPI are useful for the
CC   treatment of VD and for gene therapy
XX
SQ   Sequence 10 AA;

  Query Match             4.8%;  Score 10;  DB 4;  Length 10;
  Best Local Similarity   100.0%;  Pred. No. 0.014;
  Matches   10;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        167 FSSGITGCVK 176
              ||||||||||
Db          1 FSSGITGCVK 10

RESULT 9
ABB55909
ID   ABB55909 standard; peptide; 15 AA.
XX
AC   ABB55909;
XX
DT   15-FEB-2002  (first entry)
XX
DE   Vascular dementia-associated protein isoform (VPI) 109.
XX
KW   Vascular Dementia; VD; VD-associated protein isoform; VPI; screening;
KW   diagnosis; prognosis; gene therapy.
XX
OS   Homo sapiens.
XX
PN   WO200169261-A2.
XX
PD   20-SEP-2001.
XX
PF   14-MAR-2001; 2001WO-GB001106.
XX
PR   15-MAR-2000; 2000GB-00006285.
PR   24-NOV-2000; 2000GB-00028734.
PR   28-NOV-2000; 2000US-00724391.
XX
PA   (OXFO-) OXFORD GLYCOSCIENCES UK LTD.
XX
PI   Herath HMAC,  Parekh RB,  Rohlff C;
XX
DR   WPI; 2001-557937/62.
XX
PT   Screening, diagnosis or prognosis of vascular dementia (VD), useful for
PT   determining stage of VD and monitoring the effect of VD therapy,
PT   comprises analyzing body fluid by 2-dimensional electrophoresis for
PT   features correlated with VD.
XX
PS   Claim 6; Page 32; 151pp; English.
XX
CC   The invention relates to screening, diagnosis or prognosis of Vascular
CC   Dementia (VD) in a subject comprising analysing body fluid from the
CC   subject by 2-dimensional (2-D) electrophoresis to generate a 2-D array of
CC   features containing at least one chosen feature whose relative abundance
CC   correlates with the presence, absence, stage or severity of VD or
CC   predicts the onset or course of VD, especially detecting in a sample of
CC   cerebrospinal fluid (CSF) from the subject one of 223 VD-associated
CC   protein isoforms (VPIs) (ABB55801-ABB56295) as fully defined in the
CC   specification. Detecting VD-associated features and VPI is useful for the
CC   screening, diagnosis or prognosis of VD, for determining the stage or
CC   severity of VD, for identifying a subject at risk of VD or for monitoring
CC   the effect of therapy administered to a subject having VD. Nucleic acids
CC   encoding a VPI or inhibiting the function of a VPI are useful for the
CC   treatment of VD and for gene therapy
XX
SQ   Sequence 15 AA;

  Query Match             4.3%;  Score 9;  DB 4;  Length 15;
  Best Local Similarity   100.0%;  Pred. No. 0.2;
  Matches    9;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         92 YQLGSGEAR 100
              |||||||||
Db          1 YQLGSGEAR 9

RESULT 10
ABM65732
ID   ABM65732 standard; protein; 159 AA.
XX
AC   ABM65732;
XX
DT   20-OCT-2003  (first entry)
XX
DE   Propionibacterium acnes immunogenic polypeptide #30408.
XX
KW   Acne vulgaris; antiseborrhoeic; dermatological; antibacterial;
KW   immunostimulant; immune response; vaccine; immunogenic.
XX
OS   Propionibacterium acnes.
XX
PN   WO2003033515-A1.
XX
PD   24-APR-2003.
XX
PF   11-OCT-2002; 2002WO-US032727.
```

```
XX
PR   15-OCT-2001; 2001US-00978825.
XX
PA   (CORI-) CORIXA CORP.
XX
PI   Mitcham JL,  Skeiky YAW,  Persing DH,  Bhatia A,  Maisonneuve JL;
PI   Zhang Y,  Wang S,  Jen S,  Lodes MJ,  Benson DR,  Jones R,  Carter D;
PI   Barth B,  Vallieve-Douglass J;
XX
DR   WPI; 2003-381789/36.
XX
PT   New Propionibacterium acnes polypeptides and polynucleotides encoding the
PT   polypeptide, useful for diagnosing, preventing or treating acne vulgaris,
PT   or for stimulating an immune response specific for a P. acnes protein.
XX
PS   Claim 7; SEQ ID NO 30408; 1481pp; English.
XX
CC   The invention relates to an isolated polynucleotide (ACF64435-ACF64733)
CC   encoding a Propionibacterium acnes protein. The invention also relates to
CC   polypeptides encoded by the polynucleotides (ABM35624-ABM64536) and to
CC   immunogenic fragments of P. acnes polypeptides. The invention
CC   additionally encompasses expression vectors and host cells comprising a
CC   polynucleotide of the invention; antibodies against polypeptides of the
CC   invention; fusion proteins comprising a polypeptide of the invention; a
CC   method for stimulating an immune response specific for a P. acnes
CC   polypeptide and an isolated T cell population comprising T cells prepared
CC   via this method; a vaccine composition (comprising P. acnes polypeptides,
CC   polynucleotides, antibodies, fusion proteins, T cell populations, or
CC   antigen-presenting cells that express the polypeptide); a method and kit
CC   for detecting or determining the presence or absence of P. acnes in a
CC   patient; and a method for inhibiting the development of P. acnes in a
CC   patient. The P. acnes polypeptides, polynucleotides, antibodies, fusion
CC   proteins, T cell populations or antigen-presenting cells that express the
CC   polypeptides are useful for diagnosing, preventing or treating acne
CC   vulgaris, or for stimulating an immune response specific for a P. acnes
CC   protein. The polynucleotides can also be used as probes or primers for
CC   nucleic acid hybridisation. The vaccine composition is useful for the
CC   stimulation of an immune response against P. acnes, or for treating acne,
CC   and the kit is useful for performing a diagnostic assay. The present
CC   sequence represents a specifically claimed P. acnes polypeptide which is
CC   thought to contain an immunogenic region. Note: The sequence data for
CC   this patent did not form part of the printed specification, but was
CC   obtained in electronic format directly from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 159 AA;

  Query Match              4.3%;  Score 9;  DB 6;  Length 159;
  Best Local Similarity   100.0%;  Pred. No. 1.7;
  Matches    9;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         53 VRTSTASGL 61
              |||||||||
Db         31 VRTSTASGL 39

RESULT 11
ABB70376
ID   ABB70376 standard; protein; 3319 AA.
XX
AC   ABB70376;
XX
DT   26-MAR-2002  (first entry)
XX
DE   Drosophila melanogaster polypeptide SEQ ID NO 37920.
XX
KW   Drosophila; developmental biology; cell signalling; insecticide;
KW   pharmaceutical.
XX
OS   Drosophila melanogaster.
XX
PN   WO200171042-A2.
XX
```

```
PD   27-SEP-2001.
XX
PF   23-MAR-2001; 2001WO-US009231.
XX
PR   23-MAR-2000; 2000US-0191637P.
PR   11-JUL-2000; 2000US-00614150.
XX
PA   (PEKE ) PE CORP NY.
XX
PI   Venter JC,  Adams M,  Li PWD,  Myers EW;
XX
DR   WPI; 2001-656860/75.
DR   N-PSDB; ABL14479.
XX
PT   New isolated nucleic acid detection reagent for detecting 1000 or more
PT   genes from Drosophila and for elucidating cell signaling and cell-cell
PT   interactions.
XX
PS   Disclosure; SEQ ID NO 37920; 21pp + Sequence Listing; English.
XX
CC   The invention relates to an isolated nucleic acid detection reagent
CC   capable of detecting 1000 or more genes from Drosophila. The invention is
CC   useful in developmental biology and in elucidating cell signalling and
CC   cell-cell interactions in higher eukaryotes for the development of
CC   insecticides, therapeutics and pharmaceutical drugs. The invention
CC   discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC   sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-
CC   ABB72072). The sequence data for this patent did not form part of the
CC   printed specification, but was obtained in electronic format directly
CC   from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 3319 AA;

  Query Match              4.3%;  Score 9;  DB 4;  Length 3319;
  Best Local Similarity   100.0%;  Pred. No. 25;
  Matches    9;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        108 NDGEWHRVT 116
              |||||||||
Db       3033 NDGEWHRVT 3041

RESULT 12
ABB55907
ID   ABB55907 standard; peptide; 8 AA.
XX
AC   ABB55907;
XX
DT   15-FEB-2002  (first entry)
XX
DE   Vascular dementia-associated protein isoform (VPI) 107.
XX
KW   Vascular Dementia; VD; VD-associated protein isoform; VPI; screening;
KW   diagnosis; prognosis; gene therapy.
XX
OS   Homo sapiens.
XX
PN   WO200169261-A2.
XX
PD   20-SEP-2001.
XX
PF   14-MAR-2001; 2001WO-GB001106.
XX
PR   15-MAR-2000; 2000GB-00006285.
PR   24-NOV-2000; 2000GB-00028734.
PR   28-NOV-2000; 2000US-00724391.
XX
PA   (OXFO-) OXFORD GLYCOSCIENCES UK LTD.
XX
PI   Herath HMAC,  Parekh RB,  Rohlff C;
XX
DR   WPI; 2001-557937/62.
XX
```

```
PT   Screening, diagnosis or prognosis of vascular dementia (VD), useful for
PT   determining stage of VD and monitoring the effect of VD therapy,
PT   comprises analyzing body fluid by 2-dimensional electrophoresis for
PT   features correlated with VD.
XX
PS   Claim 6; Page 32; 151pp; English.
XX
CC   The invention relates to screening, diagnosis or prognosis of Vascular
CC   Dementia (VD) in a subject comprising analysing body fluid from the
CC   subject by 2-dimensional (2-D) electrophoresis to generate a 2-D array of
CC   features containing at least one chosen feature whose relative abundance
CC   correlates with the presence, absence, stage or severity of VD or
CC   predicts the onset or course of VD, especially detecting in a sample of
CC   cerebrospinal fluid (CSF) from the subject one of 223 VD-associated
CC   protein isoforms (VPIs) (ABB55801-ABB56295) as fully defined in the
CC   specification. Detecting VD-associated features and VPI is useful for the
CC   screening, diagnosis or prognosis of VD, for determining the stage or
CC   severity of VD, for identifying a subject at risk of VD or for monitoring
CC   the effect of therapy administered to a subject having VD. Nucleic acids
CC   encoding a VPI or inhibiting the function of a VPI are useful for the
CC   treatment of VD and for gene therapy
XX
SQ   Sequence 8 AA;

  Query Match             3.8%;  Score 8;  DB 4;  Length 8;
  Best Local Similarity   100.0%;  Pred. No. 1.4e+06;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        177 NLVLHSAR 184
              ||||||||
Db          1 NLVLHSAR 8

RESULT 13
ABB55906
ID   ABB55906 standard; peptide; 14 AA.
XX
AC   ABB55906;
XX
DT   15-FEB-2002  (first entry)
XX
DE   Vascular dementia-associated protein isoform (VPI) 106.
XX
KW   Vascular Dementia; VD; VD-associated protein isoform; VPI; screening;
KW   diagnosis; prognosis; gene therapy.
XX
OS   Homo sapiens.
XX
PN   WO200169261-A2.
XX
PD   20-SEP-2001.
XX
PF   14-MAR-2001; 2001WO-GB001106.
XX
PR   15-MAR-2000; 2000GB-00006285.
PR   24-NOV-2000; 2000GB-00028734.
PR   28-NOV-2000; 2000US-00724391.
XX
PA   (OXFO-) OXFORD GLYCOSCIENCES UK LTD.
XX
PI   Herath HMAC,  Parekh RB,  Rohlff C;
XX
DR   WPI; 2001-557937/62.
XX
PT   Screening, diagnosis or prognosis of vascular dementia (VD), useful for
PT   determining stage of VD and monitoring the effect of VD therapy,
PT   comprises analyzing body fluid by 2-dimensional electrophoresis for
PT   features correlated with VD.
XX
PS   Claim 6; Page 32; 151pp; English.
XX
CC   The invention relates to screening, diagnosis or prognosis of Vascular
CC   Dementia (VD) in a subject comprising analysing body fluid from the
```

```
CC   subject by 2-dimensional (2-D) electrophoresis to generate a 2-D array of
CC   features containing at least one chosen feature whose relative abundance
CC   correlates with the presence, absence, stage or severity of VD or
CC   predicts the onset or course of VD, especially detecting in a sample of
CC   cerebrospinal fluid (CSF) from the subject one of 223 VD-associated
CC   protein isoforms (VPIs) (ABB55801-ABB56295) as fully defined in the
CC   specification. Detecting VD-associated features and VPI is useful for the
CC   screening, diagnosis or prognosis of VD, for determining the stage or
CC   severity of VD, for identifying a subject at risk of VD or for monitoring
CC   the effect of therapy administered to a subject having VD. Nucleic acids
CC   encoding a VPI or inhibiting the function of a VPI are useful for the
CC   treatment of VD and for gene therapy
XX
SQ   Sequence 14 AA;

  Query Match             3.8%;  Score 8;  DB 4;  Length 14;
  Best Local Similarity   100.0%;  Pred. No. 1.9;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        124 GSIQVDGE 131
              ||||||||
Db          1 GSIQVDGE 8

RESULT 14
AAU40544
ID   AAU40544 standard; protein; 54 AA.
XX
AC   AAU40544;
XX
DT   13-FEB-2002  (first entry)
XX
DE   Propionibacterium acnes immunogenic protein #1440.
XX
KW   SAPHO syndrome; synovitis; acne; pustulosis; hypertosis; osteomyelitis;
KW   uveitis; endophthalmitis; bone; joint; central nervous system; ELISA;
KW   inflammatory lesion; acne vulgaris; enzyme linked immunosorbent assay;
KW   dermatological; osteopathic; neuroprotectant.
XX
OS   Propionibacterium acnes.
XX
PN   WO200181581-A2.
XX
PD   01-NOV-2001.
XX
PF   20-APR-2001; 2001WO-US012865.
XX
PR   21-APR-2000; 2000US-0199047P.
PR   02-JUN-2000; 2000US-0208841P.
PR   07-JUL-2000; 2000US-0216747P.
XX
PA   (CORI-) CORIXA CORP.
XX
PI   Skeiky YAW,  Persing DH,  Mitcham JL,  Wang SS,  Bhatia A;
PI   L'maisonneuve J,  Zhang Y,  Jen S,  Carter D;
XX
DR   WPI; 2001-616774/71.
DR   N-PSDB; AAS59512.
XX
PT   Propionibacterium acnes polypeptides and nucleic acids useful for
PT   vaccinating against and diagnosing infections, especially useful for
PT   treating acne vulgaris.
XX
PS   Example 1; SEQ ID NO 1739; 1069pp; English.
XX
CC   Sequences AAU39105-AAU68017 represent Propionibacterium acnes immunogenic
CC   polypeptides. The proteins and their associated DNA sequences are used in
CC   the treatment, prevention and diagnosis of medical conditions caused by
CC   P. acnes. The disorders include SAPHO syndrome (synovitis, acne,
CC   pustulosis, hypertosis and osteomyelitis), uveitis and endophthalmitis.
CC   P. acnes is also involved in infections of bone, joints and the central
CC   nervous system, however it is particularly involved in the inflammatory
CC   lesions associated with acne vulgaris. A method for detecting the
```

```
CC   presence or absence of P. acnes in a patient comprises contacting a
CC   sample with a binding agent that binds to the proteins of the invention
CC   and determining the amount of bound protein in the sample. The
CC   polypeptides may be used as antigens in the production of antibodies
CC   specific for P. acnes proteins. These antibodies can be used to
CC   downregulate expression and activity of P. acnes polypeptides and
CC   therefore treat P. acnes infections. The antibodies may also be used as
CC   diagnostic agents for determining P. acnes presence, for example, by
CC   enzyme linked immunosorbent assay (ELISA). Note: The sequence data for
CC   this patent did not form part of the printed specification, but was
CC   obtained in electronic format directly from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 54 AA;

  Query Match              3.8%;  Score 8;  DB 4;  Length 54;
  Best Local Similarity  100.0%;  Pred. No. 6.3;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         53 VRTSTASG 60
              ||||||||
Db         31 VRTSTASG 38

RESULT 15
ABM37063
ID   ABM37063 standard; protein; 54 AA.
XX
AC   ABM37063;
XX
DT   20-OCT-2003  (first entry)
XX
DE   Propionibacterium acnes predicted ORF-encoded polypeptide #1739.
XX
KW   Acne vulgaris; antiseborrhoeic; dermatological; antibacterial;
KW   immunostimulant; immune response; vaccine.
XX
OS   Propionibacterium acnes.
XX
PN   WO2003033515-A1.
XX
PD   24-APR-2003.
XX
PF   11-OCT-2002; 2002WO-US032727.
XX
PR   15-OCT-2001; 2001US-00978825.
XX
PA   (CORI-) CORIXA CORP.
XX
PI   Mitcham JL,  Skeiky YAW,  Persing DH,  Bhatia A,  Maisonneuve JL;
PI   Zhang Y,  Wang S,  Jen S,  Lodes MJ,  Benson DR,  Jones R,  Carter D;
PI   Barth B,  Vallieve-Douglass J;
XX
DR   WPI; 2003-381789/36.
DR   N-PSDB; ACF64441.
XX
PT   New Propionibacterium acnes polypeptides and polynucleotides encoding the
PT   polypeptide, useful for diagnosing, preventing or treating acne vulgaris,
PT   or for stimulating an immune response specific for a P. acnes protein.
XX
PS   Example 1; SEQ ID NO 1739; 1481pp; English.
XX
CC   The invention relates to an isolated polynucleotide (ACF64435-ACF64733)
CC   encoding a Propionibacterium acnes protein. The invention also relates to
CC   polypeptides encoded by the polynucleotides (ABM35624-ABM64536) and to
CC   immunogenic fragments of P. acnes polypeptides. The invention
CC   additionally encompasses expression vectors and host cells comprising a
CC   polynucleotide of the invention; antibodies against polypeptides of the
CC   invention; fusion proteins comprising a polypeptide of the invention; a
CC   method for stimulating an immune response specific for a P. acnes
CC   polypeptide and an isolated T cell population comprising T cells prepared
CC   via this method; a vaccine composition (comprising P. acnes polypeptides,
CC   polynucleotides, antibodies, fusion proteins, T cell populations, or
CC   antigen-presenting cells that express the polypeptide); a method and kit
CC   for detecting or determining the presence or absence of P. acnes in a
CC   patient; and a method for inhibiting the development of P. acnes in a
CC   patient. The P. acnes polypeptides, polynucleotides, antibodies, fusion
CC   proteins, T cell populations or antigen-presenting cells that express the
CC   polypeptides are useful for diagnosing, preventing or treating acne
CC   vulgaris, or for stimulating an immune response specific for a P. acnes
CC   protein. The polynucleotides can also be used as probes or primers for
CC   nucleic acid hybridisation. The vaccine composition is useful for the
CC   stimulation of an immune response against P. acnes, or for treating acne,
CC   and the kit is useful for performing a diagnostic assay. The present
CC   sequence represents a polypeptide predicted to be encoded by an ORF (open
CC   reading frame) contained within the P. acnes polynucleotides of the
CC   invention. Note: The sequence data for this patent did not form part of
CC   the printed specification, but was obtained in electronic format directly
CC   from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 54 AA;

  Query Match              3.8%;  Score 8;  DB 6;  Length 54;
  Best Local Similarity  100.0%;  Pred. No. 6.3;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         53 VRTSTASG 60
              ||||||||
Db         31 VRTSTASG 38

Search completed: March  9, 2004, 17:27:55
Job time : 26.5713 secs
```

```
                     GenCore version 5.1.6
           Copyright (c) 1993 - 2004  Compugen Ltd.

OM protein - protein search, using sw model

Run on:         March  9, 2004, 17:25:24 ; Search time 8.05671 Seconds
                                           (without alignments)
                                           1345.642 Million cell updates/sec

Title:          US-10-006-011A-10
Perfect score:  210
Sequence:       1 GIAESDWHLEGSGGNDAPGQ..........QPLDLQHRAQAGANTRPCPS 210

Scoring table:  OLIGO
                Gapop 60.0 , Gapext 60.0

Searched:       389414 seqs, 51625971 residues

Word size :       0

Total number of hits satisfying chosen parameters:      389414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :        Issued_Patents_AA:*
                1:  /cgn2_6/ptodata/2/iaa/5A_COMB.pep:*
                2:  /cgn2_6/ptodata/2/iaa/5B_COMB.pep:*
                3:  /cgn2_6/ptodata/2/iaa/6A_COMB.pep:*
                4:  /cgn2_6/ptodata/2/iaa/6B_COMB.pep:*
                5:  /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep:*
                6:  /cgn2_6/ptodata/2/iaa/backfiles1.pep:*

      Pred. No. is the number of results predicted by chance to have a
      score greater than or equal to the score of the result being printed,
      and is derived by analysis of the total score distribution.

                                  SUMMARIES
                     %
Result              Query
  No.      Score    Match Length DB  ID                        Description
------------------------------------------------------------------------------
    1        8      3.8   1130  2  US-08-460-309-2            Sequence 2, Appli
    2        8      3.8   1130  2  US-08-125-077-2            Sequence 2, Appli
    3        8      3.8   1130  6  5444158-2                  Patent No. 5444158
    4        8      3.8   3084  4  US-09-562-702A-12          Sequence 12, Appl
    5        8      3.8   3088  4  US-09-562-702A-8           Sequence 8, Appli
    6        8      3.8   3089  4  US-09-562-702A-4           Sequence 4, Appli
    7        8      3.8   3106  4  US-09-562-702A-10          Sequence 10, Appl
    8        8      3.8   3110  4  US-09-562-702A-2           Sequence 2, Appli
    9        8      3.8   3110  4  US-09-562-702A-6           Sequence 6, Appli
   10        8      3.8   3110  4  US-09-561-709B-7           Sequence 7, Appli
   11        8      3.8   3111  2  US-08-460-309-4            Sequence 4, Appli
   12        8      3.8   3111  2  US-08-125-077-4            Sequence 4, Appli
   13        7      3.3     77  4  US-09-252-991A-18061       Sequence 18061, A
   14        7      3.3    165  4  US-09-252-991A-27759       Sequence 27759, A
   15        7      3.3    181  4  US-09-252-991A-23483       Sequence 23483, A
   16        7      3.3    186  4  US-09-107-532A-6672        Sequence 6672, Ap
   17        7      3.3    216  1  US-08-315-695-20           Sequence 20, Appl
   18        7      3.3    237  4  US-09-252-991A-28116       Sequence 28116, A
   19        7      3.3    240  4  US-09-570-856B-16          Sequence 16, Appl
   20        7      3.3    244  4  US-09-252-991A-18465       Sequence 18465, A
   21        7      3.3    248  2  US-08-755-559-1            Sequence 1, Appli
   22        7      3.3    248  3  US-09-210-474-1            Sequence 1, Appli
   23        7      3.3    248  4  US-09-539-774-1            Sequence 1, Appli
   24        7      3.3    273  4  US-09-252-991A-28324       Sequence 28324, A
   25        7      3.3    349  4  US-09-162-524-3            Sequence 3, Appli
   26        7      3.3    352  4  US-09-252-991A-19989       Sequence 19989, A
   27        7      3.3    372  4  US-09-252-991A-20108       Sequence 20108, A
   28        7      3.3    381  3  US-08-911-853-27           Sequence 27, Appl
   29        7      3.3    381  3  US-09-479-409-27           Sequence 27, Appl
   30        7      3.3    381  4  US-09-479-453-27           Sequence 27, Appl
   31        7      3.3    384  4  US-09-252-991A-24086       Sequence 24086, A
   32        7      3.3    410  4  US-09-252-991A-31487       Sequence 31487, A
   33        7      3.3    519  4  US-09-328-352-7681         Sequence 7681, Ap
   34        7      3.3    728  4  US-09-107-532A-4518        Sequence 4518, Ap
   35        7      3.3    780  4  US-09-252-991A-19120       Sequence 19120, A
   36        7      3.3    814  4  US-09-486-072-1            Sequence 1, Appli
   37        7      3.3    881  4  US-09-486-072-2            Sequence 2, Appli
   38        7      3.3    915  4  US-09-489-039A-8249        Sequence 8249, Ap
   39        7      3.3   1094  4  US-09-268-347-32           Sequence 32, Appl
   40        7      3.3   1940  2  US-08-644-271-30           Sequence 30, Appl
   41        7      3.3   1940  4  US-09-077-955-34           Sequence 34, Appl
   42        7      3.3   2175  4  US-09-404-650-2            Sequence 2, Appli
   43        7      3.3   2175  4  US-09-935-541-2            Sequence 2, Appli
   44        7      3.3   2188  4  US-09-404-650-4            Sequence 4, Appli
   45        7      3.3   2188  4  US-09-935-541-4            Sequence 4, Appli

                                  ALIGNMENTS

RESULT 1
US-08-460-309-2
; Sequence 2, Application US/08460309
; Patent No. 5837496
;  GENERAL INFORMATION:
;    APPLICANT:  Engvall, Eva
;    APPLICANT:  Leivo, Ilmo
;    TITLE OF INVENTION:  Nucleic Acids Encoding Merosin, Merosin
;    TITLE OF INVENTION:  Fragments and Uses Thereof
;    NUMBER OF SEQUENCES:  23
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Campbell and Flores
;      STREET:  4370 La Jolla Village Drive, Suite 700
;      CITY:  San Diego
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  92122
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/460,309
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/125,077
;      FILING DATE:  22-SEP-1993
;      APPLICATION NUMBER:  US PCT/US 94/10730
;      FILING DATE:  21-SEP-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/472,319
;      FILING DATE:  30-JAN-1990
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/919,951
;      FILING DATE:  27-JUL-1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Campbell, Cathryn A.
;      REGISTRATION NUMBER:  31,815
;      REFERENCE/DOCKET NUMBER:  P-LA 9721
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (619) 535-9001
;      TELEFAX:  (619) 535-8949
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1130 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
```

```
US-08-460-309-2

  Query Match             3.8%;  Score 8;  DB 2;  Length 1130;
  Best Local Similarity   100.0%;  Pred. No. 24;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         48 TIELEVRT 55
              ||||||||
Db        808 TIELEVRT 815


RESULT 2
US-08-125-077-2
; Sequence 2, Application US/08125077
; Patent No. 5872231
; Patent No. 5872231 5840863
;  GENERAL INFORMATION:
;    APPLICANT:  Engvall, Eva
;    APPLICANT:  Leivo, Ilmo
;    TITLE OF INVENTION:  Nucleic Acids Encoding Merosin, Merosin
;    TITLE OF INVENTION:  Fragments and Uses Thereof
;    NUMBER OF SEQUENCES:  23
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Campbell and Flores
;      STREET:  4370 La Jolla Village Drive, Suite 700
;      CITY:  San Diego
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  92122
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/125,077
;      FILING DATE:  22-SEP-1993
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US PCT/US 94/10730
;      FILING DATE:  21-SEP-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/472,319
;      FILING DATE:  30-JAN-1990
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/919,951
;      FILING DATE:  27-JUL-1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Campbell, Cathryn A.
;      REGISTRATION NUMBER:  31,815
;      REFERENCE/DOCKET NUMBER:  P-LA 9721
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (619) 535-9001
;      TELEFAX:  (619) 535-8949
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1130 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-125-077-2

  Query Match             3.8%;  Score 8;  DB 2;  Length 1130;
  Best Local Similarity   100.0%;  Pred. No. 24;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         48 TIELEVRT 55
              ||||||||
Db        808 TIELEVRT 815


RESULT 3
5444158-2
;Patent No. 5444158
;    APPLICANT: ENGVALL, EVA;SANES, JOSHUA
;    TITLE OF INVENTION: MEROSIN, NUCLEIC ACIDS ENCODING,
;FRAGMENTS AND USES THEREOF
;    NUMBER OF SEQUENCES: 4
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/87,642
;      FILING DATE: 08-JUL-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER: 587,689
;      FILING DATE: 24-SEP-1990
;      APPLICATION NUMBER: 472,319
;      FILING DATE: 30-JAN-1990
;SEQ ID NO:2:
;      LENGTH: 1130
5444158-2

  Query Match             3.8%;  Score 8;  DB 6;  Length 1130;
  Best Local Similarity   100.0%;  Pred. No. 24;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         48 TIELEVRT 55
              ||||||||
Db        808 TIELEVRT 815


RESULT 4
US-09-562-702A-12
; Sequence 12, Application US/09562702A
; Patent No. 6632790
; GENERAL INFORMATION:
;  APPLICANT: Yurchenco, Peter
;  TITLE OF INVENTION: Laminin 2 and Methods for Its Use
;  FILE REFERENCE: 99-274-B
;  CURRENT APPLICATION NUMBER: US/09/562,702A
;  CURRENT FILING DATE:  2000-04-28
;  PRIOR APPLICATION NUMBER: 60/155,945
;  PRIOR FILING DATE: 1999-09-24
;  PRIOR APPLICATION NUMBER: 60/143,289
;  PRIOR FILING DATE: 1999-07-12
;  PRIOR APPLICATION NUMBER: 60/139,198
;  PRIOR FILING DATE: 1999-06-15
;  PRIOR APPLICATION NUMBER: 60/131,720
;  PRIOR FILING DATE: 1999-04-30
;  NUMBER OF SEQ ID NOS: 32
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 12
;   LENGTH: 3084
;   TYPE: PRT
;   ORGANISM: Mus musculus
US-09-562-702A-12

  Query Match             3.8%;  Score 8;  DB 4;  Length 3084;
  Best Local Similarity   100.0%;  Pred. No. 61;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         48 TIELEVRT 55
              ||||||||
Db       2762 TIELEVRT 2769


RESULT 5
US-09-562-702A-8
; Sequence 8, Application US/09562702A
; Patent No. 6632790
; GENERAL INFORMATION:
;  APPLICANT: Yurchenco, Peter
;  TITLE OF INVENTION: Laminin 2 and Methods for Its Use
;  FILE REFERENCE: 99-274-B
;  CURRENT APPLICATION NUMBER: US/09/562,702A
;  CURRENT FILING DATE:  2000-04-28
;  PRIOR APPLICATION NUMBER: 60/155,945
;  PRIOR FILING DATE: 1999-09-24
```

```
;  PRIOR APPLICATION NUMBER: 60/143,289
;  PRIOR FILING DATE: 1999-07-12
;  PRIOR APPLICATION NUMBER: 60/139,198
;  PRIOR FILING DATE: 1999-06-15
;  PRIOR APPLICATION NUMBER: 60/131,720
;  PRIOR FILING DATE: 1999-04-30
;  NUMBER OF SEQ ID NOS: 32
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 8
;   LENGTH: 3088
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-562-702A-8

  Query Match             3.8%;  Score 8;  DB 4;  Length 3088;
  Best Local Similarity   100.0%;  Pred. No. 61;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         48 TIELEVRT 55
              ||||||||
Db       2766 TIELEVRT 2773

RESULT 6
US-09-562-702A-4
; Sequence 4, Application US/09562702A
; Patent No. 6632790
; GENERAL INFORMATION:
;  APPLICANT: Yurchenco, Peter
;  TITLE OF INVENTION: Laminin 2 and Methods for Its Use
;  FILE REFERENCE: 99-274-B
;  CURRENT APPLICATION NUMBER: US/09/562,702A
;  CURRENT FILING DATE:  2000-04-28
;  PRIOR APPLICATION NUMBER: 60/155,945
;  PRIOR FILING DATE: 1999-09-24
;  PRIOR APPLICATION NUMBER: 60/143,289
;  PRIOR FILING DATE: 1999-07-12
;  PRIOR APPLICATION NUMBER: 60/139,198
;  PRIOR FILING DATE: 1999-06-15
;  PRIOR APPLICATION NUMBER: 60/131,720
;  PRIOR FILING DATE: 1999-04-30
;  NUMBER OF SEQ ID NOS: 32
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
;   LENGTH: 3089
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-562-702A-4

  Query Match             3.8%;  Score 8;  DB 4;  Length 3089;
  Best Local Similarity   100.0%;  Pred. No. 61;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         48 TIELEVRT 55
              ||||||||
Db       2766 TIELEVRT 2773

RESULT 7
US-09-562-702A-10
; Sequence 10, Application US/09562702A
; Patent No. 6632790
; GENERAL INFORMATION:
;  APPLICANT: Yurchenco, Peter
;  TITLE OF INVENTION: Laminin 2 and Methods for Its Use
;  FILE REFERENCE: 99-274-B
;  CURRENT APPLICATION NUMBER: US/09/562,702A
;  CURRENT FILING DATE:  2000-04-28
;  PRIOR APPLICATION NUMBER: 60/155,945
;  PRIOR FILING DATE: 1999-09-24
;  PRIOR APPLICATION NUMBER: 60/143,289
;  PRIOR FILING DATE: 1999-07-12
;  PRIOR APPLICATION NUMBER: 60/139,198
;  PRIOR FILING DATE: 1999-06-15
;  PRIOR APPLICATION NUMBER: 60/131,720
;  PRIOR FILING DATE: 1999-04-30
;  NUMBER OF SEQ ID NOS: 32
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 10
;   LENGTH: 3106
;   TYPE: PRT
;   ORGANISM: Mus musculus
US-09-562-702A-10

  Query Match             3.8%;  Score 8;  DB 4;  Length 3106;
  Best Local Similarity   100.0%;  Pred. No. 61;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         48 TIELEVRT 55
              ||||||||
Db       2784 TIELEVRT 2791

RESULT 8
US-09-562-702A-2
; Sequence 2, Application US/09562702A
; Patent No. 6632790
; GENERAL INFORMATION:
;  APPLICANT: Yurchenco, Peter
;  TITLE OF INVENTION: Laminin 2 and Methods for Its Use
;  FILE REFERENCE: 99-274-B
;  CURRENT APPLICATION NUMBER: US/09/562,702A
;  CURRENT FILING DATE:  2000-04-28
;  PRIOR APPLICATION NUMBER: 60/155,945
;  PRIOR FILING DATE: 1999-09-24
;  PRIOR APPLICATION NUMBER: 60/143,289
;  PRIOR FILING DATE: 1999-07-12
;  PRIOR APPLICATION NUMBER: 60/139,198
;  PRIOR FILING DATE: 1999-06-15
;  PRIOR APPLICATION NUMBER: 60/131,720
;  PRIOR FILING DATE: 1999-04-30
;  NUMBER OF SEQ ID NOS: 32
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
;   LENGTH: 3110
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-562-702A-2

  Query Match             3.8%;  Score 8;  DB 4;  Length 3110;
  Best Local Similarity   100.0%;  Pred. No. 61;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         48 TIELEVRT 55
              ||||||||
Db       2788 TIELEVRT 2795

RESULT 9
US-09-562-702A-6
; Sequence 6, Application US/09562702A
; Patent No. 6632790
; GENERAL INFORMATION:
;  APPLICANT: Yurchenco, Peter
;  TITLE OF INVENTION: Laminin 2 and Methods for Its Use
;  FILE REFERENCE: 99-274-B
;  CURRENT APPLICATION NUMBER: US/09/562,702A
;  CURRENT FILING DATE:  2000-04-28
;  PRIOR APPLICATION NUMBER: 60/155,945
;  PRIOR FILING DATE: 1999-09-24
;  PRIOR APPLICATION NUMBER: 60/143,289
;  PRIOR FILING DATE: 1999-07-12
;  PRIOR APPLICATION NUMBER: 60/139,198
;  PRIOR FILING DATE: 1999-06-15
;  PRIOR APPLICATION NUMBER: 60/131,720
;  PRIOR FILING DATE: 1999-04-30
```

```
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 6
;   LENGTH: 3110
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-562-702A-6

  Query Match             3.8%;  Score 8;  DB 4;  Length 3110;
  Best Local Similarity   100.0%;  Pred. No. 61;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         48 TIELEVRT 55
              ||||||||
Db       2788 TIELEVRT 2795


RESULT 10
US-09-561-709B-7
; Sequence 7, Application US/09561709B
; Patent No. 6682911
; GENERAL INFORMATION:
;  APPLICANT: Burgeson, Robert
;  APPLICANT:  Champliaud, Marie-France
;  APPLICANT:  Olson, Pamela
;  APPLICANT:  Koch, Manuel
;  APPLICANT:  Brunken, William
;  TITLE OF INVENTION: LAMININS AND USES THEREOF
;  FILE REFERENCE: 10287-060001
;  CURRENT APPLICATION NUMBER: US/09/561,709B
;  CURRENT FILING DATE:  2000-05-01
;  PRIOR APPLICATION NUMBER: US 09/168,949
;  PRIOR FILING DATE: 1998-10-09
;  PRIOR APPLICATION NUMBER: US 60/061,609
;  PRIOR FILING DATE: 1997-10-10
;  NUMBER OF SEQ ID NOS: 13
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 7
;   LENGTH: 3110
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-561-709B-7

  Query Match             3.8%;  Score 8;  DB 4;  Length 3110;
  Best Local Similarity   100.0%;  Pred. No. 61;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         48 TIELEVRT 55
              ||||||||
Db       2788 TIELEVRT 2795


RESULT 11
US-08-460-309-4
; Sequence 4, Application US/08460309
; Patent No. 5837496
;  GENERAL INFORMATION:
;    APPLICANT:  Engvall, Eva
;    APPLICANT:  Leivo, Ilmo
;    TITLE OF INVENTION:  Nucleic Acids Encoding Merosin, Merosin
;    TITLE OF INVENTION:  Fragments and Uses Thereof
;    NUMBER OF SEQUENCES:  23
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Campbell and Flores
;      STREET:  4370 La Jolla Village Drive, Suite 700
;      CITY:  San Diego
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  92122
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
```

```
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/460,309
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/125,077
;      FILING DATE:  22-SEP-1993
;      APPLICATION NUMBER:  US PCT/US 94/10730
;      FILING DATE:  21-SEP-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/472,319
;      FILING DATE:  30-JAN-1990
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/919,951
;      FILING DATE:  27-JUL-1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Campbell, Cathryn A.
;      REGISTRATION NUMBER:  31,815
;      REFERENCE/DOCKET NUMBER:  P-LA 9721
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (619) 535-9001
;      TELEFAX:  (619) 535-8949
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3111 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
US-08-460-309-4

  Query Match             3.8%;  Score 8;  DB 2;  Length 3111;
  Best Local Similarity   100.0%;  Pred. No. 61;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         48 TIELEVRT 55
              ||||||||
Db       2789 TIELEVRT 2796


RESULT 12
US-08-125-077-4
; Sequence 4, Application US/08125077
; Patent No. 5872231
; Patent No. 5872231 5840863
;  GENERAL INFORMATION:
;    APPLICANT:  Engvall, Eva
;    APPLICANT:  Leivo, Ilmo
;    TITLE OF INVENTION:  Nucleic Acids Encoding Merosin, Merosin
;    TITLE OF INVENTION:  Fragments and Uses Thereof
;    NUMBER OF SEQUENCES:  23
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Campbell and Flores
;      STREET:  4370 La Jolla Village Drive, Suite 700
;      CITY:  San Diego
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  92122
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/125,077
;      FILING DATE:  22-SEP-1993
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US PCT/US 94/10730
;      FILING DATE:  21-SEP-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/472,319
;      FILING DATE:  30-JAN-1990
;    PRIOR APPLICATION DATA:
```

```
;      APPLICATION NUMBER:  US 07/919,951.
;      FILING DATE:  27-JUL-1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Campbell, Cathryn A.
;      REGISTRATION NUMBER:  31,815
;      REFERENCE/DOCKET NUMBER:  P-LA 9721
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (619) 535-9001
;      TELEFAX:  (619) 535-8949
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3111 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
US-08-125-077-4

  Query Match              3.8%;  Score 8;  DB 2;  Length 3111;
  Best Local Similarity   100.0%;  Pred. No. 61;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         48 TIELEVRT 55
              ||||||||
Db       2789 TIELEVRT 2796

RESULT 13
US-09-252-991A-18061
; Sequence 18061, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
;  APPLICANT:  Marc J. Rubenfield et al.
;  TITLE OF INVENTION:  NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
;  TITLE OF INVENTION:  AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE:  107196.136
;  CURRENT APPLICATION NUMBER: US/09/252,991A
;  CURRENT FILING DATE:  1999-02-18
;  PRIOR APPLICATION NUMBER:  US 60/074,788
;  PRIOR FILING DATE:  1998-02-18
;  PRIOR APPLICATION NUMBER:  US 60/094,190
;  PRIOR FILING DATE:  1998-07-27
;  NUMBER OF SEQ ID NOS:  33142
; SEQ ID NO 18061
;   LENGTH: 77
;   TYPE: PRT
;   ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-18061

  Query Match              3.3%;  Score 7;  DB 4;  Length 77;
  Best Local Similarity   100.0%;  Pred. No. 19;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        182 SARPGAP 188
              |||||||
Db         49 SARPGAP 55

RESULT 14
US-09-252-991A-27759
; Sequence 27759, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
;  APPLICANT:  Marc J. Rubenfield et al.
;  TITLE OF INVENTION:  NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
;  TITLE OF INVENTION:  AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE:  107196.136
;  CURRENT APPLICATION NUMBER: US/09/252,991A
;  CURRENT FILING DATE:  1999-02-18
;  PRIOR APPLICATION NUMBER:  US 60/074,788
;  PRIOR FILING DATE:  1998-02-18
;  PRIOR APPLICATION NUMBER:  US 60/094,190
;  PRIOR FILING DATE:  1998-07-27
;  NUMBER OF SEQ ID NOS:  33142
; SEQ ID NO 27759
;   LENGTH: 165
;   TYPE: PRT
;   ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-27759

  Query Match              3.3%;  Score 7;  DB 4;  Length 165;
  Best Local Similarity   100.0%;  Pred. No. 39;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        135 SGRSPGP 141
              |||||||
Db         92 SGRSPGP 98

RESULT 15
US-09-252-991A-23483
; Sequence 23483, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
;  APPLICANT:  Marc J. Rubenfield et al.
;  TITLE OF INVENTION:  NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
;  TITLE OF INVENTION:  AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE:  107196.136
;  CURRENT APPLICATION NUMBER: US/09/252,991A
;  CURRENT FILING DATE:  1999-02-18
;  PRIOR APPLICATION NUMBER:  US 60/074,788
;  PRIOR FILING DATE:  1998-02-18
;  PRIOR APPLICATION NUMBER:  US 60/094,190
;  PRIOR FILING DATE:  1998-07-27
;  NUMBER OF SEQ ID NOS:  33142
; SEQ ID NO 23483
;   LENGTH: 181
;   TYPE: PRT
;   ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-23483

  Query Match              3.3%;  Score 7;  DB 4;  Length 181;
  Best Local Similarity   100.0%;  Pred. No. 43;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        135 SGRSPGP 141
              |||||||
Db        147 SGRSPGP 153

Search completed: March  9, 2004, 17:32:25
Job time : 9.05671 secs
```

```
                    GenCore version 5.1.6
          Copyright (c) 1993 - 2004  Compugen Ltd.

OM protein - protein search, using sw model

Run on:         March  9, 2004, 17:27:59 ; Search time 15.4128 Seconds
                                          (without alignments)
                                          2876.963 Million cell updates/sec

Title:          US-10-006-011A-10
Perfect score:  210
Sequence:       1 GIAESDWHLEGSGGNDAPGQ..........QPLDLQHRAQAGANTRPCPS 210

Scoring table:  OLIGO
                Gapop 60.0 , Gapext 60.0

Searched:       809742 seqs, 211153259 residues

Word size :      0

Total number of hits satisfying chosen parameters:      809742

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :       Published_Applications_AA:*
                  1:  /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep:*
                  2:  /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep:*
                  3:  /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep:*
                  4:  /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep:*
                  5:  /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep:*
                  6:  /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep:*
                  7:  /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep:*
                  8:  /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep:*
                  9:  /cgn2_6/ptodata/1/pubpaa/US09A_PUBCOMB.pep:*
                  10:  /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pep:*
                  11:  /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep:*
                  12:  /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep:*
                  13:  /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep:*
                  14:  /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep:*
                  15:  /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep:*
                  16:  /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep:*
                  17:  /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep:*
                  18:  /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep:*

      Pred. No. is the number of results predicted by chance to have a
      score greater than or equal to the score of the result being printed,
      and is derived by analysis of the total score distribution.

                                   SUMMARIES
                     %
Result               Query
  No.     Score      Match Length DB  ID                     Description
-------------------------------------------------------------------------------
    1       8     3.8      80  9  US-09-864-761-35784       Sequence 35784, A
    2       8     3.8    3070  10  US-09-961-403-7           Sequence 7, Appli
    3       7     3.3      29  14  US-10-029-386-32069       Sequence 32069, A
    4       7     3.3      50  9  US-09-975-143-1           Sequence 1, Appli
    5       7     3.3      61  14  US-10-074-475-177         Sequence 177, App
    6       7     3.3      72  14  US-10-106-698-6518        Sequence 6518, Ap
    7       7     3.3      91  14  US-10-156-761-8320        Sequence 8320, Ap
    8       7     3.3     122  15  US-10-108-260A-3535       Sequence 3535, Ap
    9       7     3.3     152  15  US-10-104-047-3338        Sequence 3338, Ap
   10       7     3.3     157  14  US-10-156-761-11984       Sequence 11984, A
   11       7     3.3     193  9  US-09-809-545A-4          Sequence 4, Appli
   12       7     3.3     194  9  US-09-764-870-537         Sequence 537, App
   13       7     3.3     194  14  US-10-125-540-537         Sequence 537, App
   14       7     3.3     210  9  US-09-764-870-432         Sequence 432, App
   15       7     3.3     210  14  US-10-125-540-432         Sequence 432, App
   16       7     3.3     238  9  US-09-764-853-758         Sequence 758, App
   17       7     3.3     238  9  US-09-764-898-252         Sequence 252, App
   18       7     3.3     238  10  US-09-764-881-158         Sequence 158, App
   19       7     3.3     238  14  US-10-073-865-121         Sequence 121, App
   20       7     3.3     238  15  US-10-242-747-158         Sequence 158, App
   21       7     3.3     240  14  US-10-237-386-38          Sequence 38, Appl
   22       7     3.3     245  9  US-09-738-626-5347        Sequence 5347, Ap
   23       7     3.3     248  9  US-09-799-777-32          Sequence 32, Appl
   24       7     3.3     248  9  US-09-997-165-4           Sequence 4, Appli
   25       7     3.3     248  14  US-10-080-522-1           Sequence 1, Appli
   26       7     3.3     279  14  US-10-017-161-1004        Sequence 1004, Ap
   27       7     3.3     279  15  US-10-292-798-854         Sequence 854, App
   28       7     3.3     303  15  US-10-369-493-19328       Sequence 19328, A
   29       7     3.3     324  14  US-10-281-024-15          Sequence 15, Appl
   30       7     3.3     326  14  US-10-245-538-4           Sequence 4, Appli
   31       7     3.3     349  10  US-09-820-598-3           Sequence 3, Appli
   32       7     3.3     349  13  US-10-095-932-3           Sequence 3, Appli
   33       7     3.3     375  9  US-09-978-249-10          Sequence 10, Appl
   34       7     3.3     379  9  US-09-764-853-621         Sequence 621, App
   35       7     3.3     387  14  US-10-017-161-2352        Sequence 2352, Ap
   36       7     3.3     387  15  US-10-292-798-1996        Sequence 1996, Ap
   37       7     3.3     389  15  US-10-108-260A-3014       Sequence 3014, Ap
   38       7     3.3     420  15  US-10-369-493-9782        Sequence 9782, Ap
   39       7     3.3     432  9  US-09-764-853-541         Sequence 541, App
   40       7     3.3     432  9  US-09-764-898-179         Sequence 179, App
   41       7     3.3     432  10  US-09-764-881-102         Sequence 102, App
   42       7     3.3     432  14  US-10-073-865-78          Sequence 78, Appl
   43       7     3.3     432  15  US-10-242-747-102         Sequence 102, App
   44       7     3.3     455  15  US-10-369-493-13712       Sequence 13712, A
   45       7     3.3     463  15  US-10-104-047-3058        Sequence 3058, Ap

                                 ALIGNMENTS


RESULT 1
US-09-864-761-35784
; Sequence 35784, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
;  APPLICANT: Penn, Sharron G.
;  APPLICANT:  Rank, David R.
;  APPLICANT:  Hanzel, David K.
;  APPLICANT:  Chen, Wensheng
;  TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
;  TITLE OF INVENTION:  GENE EXPRESSION ANALYSIS BY MICROARRAY
;  FILE REFERENCE: Aeomica-X-1
;  CURRENT APPLICATION NUMBER: US/09/864,761
;  CURRENT FILING DATE:  2001-05-23
;  PRIOR APPLICATION NUMBER: US 60/180,312
;  PRIOR FILING DATE: 2000-02-04
;  PRIOR APPLICATION NUMBER: US 60/207,456
;  PRIOR FILING DATE: 2000-05-26
;  PRIOR APPLICATION NUMBER: US 09/632,366
;  PRIOR FILING DATE: 2000-08-03
;  PRIOR APPLICATION NUMBER: GB 24263.6
;  PRIOR FILING DATE: 2000-10-04
;  PRIOR APPLICATION NUMBER: US 60/236,359
;  PRIOR FILING DATE: 2000-09-27
;  PRIOR APPLICATION NUMBER: PCT/US01/00666
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00667
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00664
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00669
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00665
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00668
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00663
;  PRIOR FILING DATE: 2001-01-30
```

```
;  PRIOR APPLICATION NUMBER: PCT/US01/00662
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00661
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00670
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: US 60/234,687
;  PRIOR FILING DATE: 2000-09-21
;  PRIOR APPLICATION NUMBER: US 09/608,408
;  PRIOR FILING DATE: 2000-06-30
;  PRIOR APPLICATION NUMBER: US 09/774,203
;  PRIOR FILING DATE: 2001-01-29
;  NUMBER OF SEQ ID NOS: 49117
;  SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 35784
;   LENGTH: 80
;   TYPE: PRT
;   ORGANISM: Homo sapiens
;   FEATURE:
;   OTHER INFORMATION: MAP TO AL035413.19
;   OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 4.1
;   OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 4.4
;   OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 3.2
;   OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 4.8
;   OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 6.5
;   OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL =  13
;   OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 5.8
;   OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 5.2
;   OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 6.2
;   OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL =   3
;   OTHER INFORMATION: EST_HUMAN HIT: BE875511.1, EVALUE 3.00e-14
US-09-864-761-35784

  Query Match             3.8%;  Score 8;  DB 9;  Length 80;
  Best Local Similarity   100.0%;  Pred. No. 8.6;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        185 PGAPPPQP 192
              ||||||||
Db         33 PGAPPPQP 40

RESULT 2
US-09-961-403-7
; Sequence 7, Application US/09961403
; Publication No. US20030077589A1
; GENERAL INFORMATION:
;  APPLICANT: HE-STUMPP, HOLGER
;  APPLICANT:  HAENDLER, BERNARD
;  APPLICANT:  KRAETZSCHMAR, JOERN
;  APPLICANT:  KREFT, BERTHOLT
;  APPLICANT:  WINTERHAGER, ELKE
;  APPLICANT:  REGIDOR, PEDRO
;  APPLICANT:  SCOTTI, SIMONE
;  TITLE OF INVENTION: METHOD FOR IN VITRO DIAGNOSIS OF ENDOMETRIOSIS
;  FILE REFERENCE: SCH-1789
;  CURRENT APPLICATION NUMBER: US/09/961,403
;  CURRENT FILING DATE:  2001-09-25
;  NUMBER OF SEQ ID NOS: 15
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 7
;   LENGTH: 3070
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-961-403-7

  Query Match             3.8%;  Score 8;  DB 10;  Length 3070;
  Best Local Similarity   100.0%;  Pred. No. 2.1e+02;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         48 TIELEVRT 55
              ||||||||
Db       2748 TIELEVRT 2755
```

```
RESULT 3
US-10-029-386-32069
; Sequence 32069, Application US/10029386
; Publication No. US20030194704A1
; GENERAL INFORMATION:
;  APPLICANT: Penn, Sharron G.
;  APPLICANT:  Rank, David R.
;  APPLICANT:  Hanzel, David K.
;  TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR G
;  TITLE OF INVENTION:  EXPRESSION ANALYSIS TWO
;  FILE REFERENCE: AEOMICA-X-2
;  CURRENT APPLICATION NUMBER: US/10/029,386
;  CURRENT FILING DATE:  2001-12-20
;  NUMBER OF SEQ ID NOS: 34288
;  SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 32069
;   LENGTH: 29
;   TYPE: PRT
;   ORGANISM: Homo sapiens
;   FEATURE:
;   OTHER INFORMATION: MAP TO AL022319.2
;   OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 2.1
;   OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 2.5
;   OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 4.4
;   OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 2.5
;   OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 2.3
;   OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 2.7
US-10-029-386-32069

  Query Match             3.3%;  Score 7;  DB 14;  Length 29;
  Best Local Similarity   100.0%;  Pred. No. 32;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        187 APPPQPL 193
              |||||||
Db          8 APPPQPL 14

RESULT 4
US-09-975-143-1
; Sequence 1, Application US/09975143
; Patent No. US20020155513A1
; GENERAL INFORMATION:
;  APPLICANT: HSU, Daniel, K.
;  APPLICANT:  LIU, Fu-Tong
;  APPLICANT:  DOWLING, Christopher, A.
;  TITLE OF INVENTION: GALECTIN EXPRESSION IS INDUCED IN
;  TITLE OF INVENTION:  CIRRHOTIC LIVER AND HEPATOCELLULAR CARCINOMA
;  FILE REFERENCE: DANHSU.001C1
;  CURRENT APPLICATION NUMBER: US/09/975,143
;  CURRENT FILING DATE:  2001-10-10
;  PRIOR APPLICATION NUMBER: PCT/US00/08561
;  PRIOR FILING DATE: 2000-03-29
;  NUMBER OF SEQ ID NOS: 47
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1
;   LENGTH: 50
;   TYPE: PRT
;   ORGANISM: chicken
US-09-975-143-1

  Query Match             3.3%;  Score 7;  DB 9;  Length 50;
  Best Local Similarity   100.0%;  Pred. No. 51;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        185 PGAPPPQ 191
              |||||||
Db         37 PGAPPPQ 43

RESULT 5
```

```
US-10-074-475-177
; Sequence 177, Application US/10074475
; Publication No. US20030092898A1
; GENERAL INFORMATION:
;  APPLICANT: Salceda, Susana
;  APPLICANT:  Macina, Roberto
;  APPLICANT:  Hu, Ping
;  APPLICANT:  Recipon, Herve
;  APPLICANT:  Karra, Kalpana
;  APPLICANT:  Cafferkey, Robert
;  APPLICANT:  Sun, Yongming
;  APPLICANT:  Liu, Chenghua
;  TITLE OF INVENTION: Compositions and Methods Relating to Breast Specific
;  TITLE OF INVENTION:  Genes and Proteins
;  FILE REFERENCE: DEX-0313
;  CURRENT APPLICATION NUMBER: US/10/074,475
;  CURRENT FILING DATE:  2002-02-13
;  PRIOR APPLICATION NUMBER: 60/268,292
;  PRIOR FILING DATE: 2001-02-13
;  NUMBER OF SEQ ID NOS: 295
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 177
;   LENGTH: 61
;   TYPE: PRT
;   ORGANISM: Homo sapien
US-10-074-475-177

  Query Match             3.3%;  Score 7;  DB 14;  Length 61;
  Best Local Similarity   100.0%;  Pred. No. 61;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         41 SLPEVPE 47
              |||||||
Db         22 SLPEVPE 28

RESULT 6
US-10-106-698-6518
; Sequence 6518, Application US/10106698
; Publication No. US20030109690A1
; GENERAL INFORMATION:
;  APPLICANT: Ruben et al.
;  TITLE OF INVENTION: Colon and Colon Cancer Associated Polynucleotides and Polypeptide
;  FILE REFERENCE: PA005P1
;  CURRENT APPLICATION NUMBER: US/10/106,698
;  CURRENT FILING DATE:  2002-03-27
;  PRIOR APPLICATION NUMBER: PCT/US00/26524
;  PRIOR FILING DATE: 2000-09-28
;  PRIOR APPLICATION NUMBER: US 60/157,137
;  PRIOR FILING DATE: 1999-09-29
;  PRIOR APPLICATION NUMBER: US 60/163,280
;  PRIOR FILING DATE: 1999-11-03
;  NUMBER OF SEQ ID NOS: 8564
;  SOFTWARE: PatentIn Ver. 3.0
; SEQ ID NO 6518
;   LENGTH: 72
;   TYPE: PRT
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: MISC_FEATURE
;   LOCATION: (13)
;   OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
;   NAME/KEY: MISC_FEATURE
;   LOCATION: (63)
;   OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-10-106-698-6518

  Query Match             3.3%;  Score 7;  DB 14;  Length 72;
  Best Local Similarity   100.0%;  Pred. No. 70;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        118 LREGRRG 124
              |||||||
Db         45 LREGRRG 51

RESULT 7
US-10-156-761-8320
; Sequence 8320, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
;  APPLICANT: OMURA, SATOSHI
;  APPLICANT:  IKEDA, HARUO
;  APPLICANT:  ISHIKAWA, JUN
;  APPLICANT:  HORIKAWA, HIROSHI
;  APPLICANT:  SHIBA, TADAYOSHI
;  APPLICANT:  SAKAKI, YOSHIYUKI
;  APPLICANT:  HATTORI, MASAHIRA
;  TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
;  FILE REFERENCE: 249-262
;  CURRENT APPLICATION NUMBER: US/10/156,761
;  CURRENT FILING DATE:  2002-05-29
;  PRIOR APPLICATION NUMBER: JP 2001-204089
;  PRIOR FILING DATE: 2001-05-30
;  PRIOR APPLICATION NUMBER: JP 2001-272697
;  PRIOR FILING DATE: 2001-08-02
;  NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 8320
;   LENGTH: 91
;   TYPE: PRT
;   ORGANISM: Streptomyces avermitilis
US-10-156-761-8320

  Query Match             3.3%;  Score 7;  DB 14;  Length 91;
  Best Local Similarity   100.0%;  Pred. No. 87;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        119 REGRRGS 125
              |||||||
Db         67 REGRRGS 73

RESULT 8
US-10-108-260A-3535
; Sequence 3535, Application US/10108260A
; Publication No. US20040005560A1
; GENERAL INFORMATION:
;  APPLICANT: HELIX RESEARCH INSTITUTE
;  TITLE OF INVENTION: No. US20040005560A1el full length cDNA
;  FILE REFERENCE: H1-A0106
;  CURRENT APPLICATION NUMBER: US/10/108,260A
;  CURRENT FILING DATE:  2002-03-27
;  NUMBER OF SEQ ID NOS: 5458
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3535
;   LENGTH: 122
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-108-260A-3535

  Query Match             3.3%;  Score 7;  DB 15;  Length 122;
  Best Local Similarity   100.0%;  Pred. No. 1.1e+02;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        186 GAPPPQP 192
              |||||||
Db         23 GAPPPQP 29

RESULT 9
US-10-104-047-3338
; Sequence 3338, Application US/10104047
; Publication No. US20030236392A1
; GENERAL INFORMATION:
;  APPLICANT: HELIX RESEARCH INSTITUTE
;  TITLE OF INVENTION: No. US20030236392A1el full length cDNA
```

```
;  FILE REFERENCE: H1-A0105
;  CURRENT APPLICATION NUMBER: US/10/104,047
;  CURRENT FILING DATE:  2002-03-25
;  PRIOR APPLICATION NUMBER:
;  PRIOR FILING DATE:
;  NUMBER OF SEQ ID NOS: 4096
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3338
;   LENGTH: 152
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-104-047-3338

  Query Match              3.3%;  Score 7;  DB 15;  Length 152;
  Best Local Similarity   100.0%;  Pred. No. 1.4e+02;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         77 DFISLGL 83
              |||||||
Db         25 DFISLGL 31


RESULT 10
US-10-156-761-11984
; Sequence 11984, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
;  APPLICANT: OMURA, SATOSHI
;  APPLICANT:  IKEDA, HARUO
;  APPLICANT:  ISHIKAWA, JUN
;  APPLICANT:  HORIKAWA, HIROSHI
;  APPLICANT:  SHIBA, TADAYOSHI
;  APPLICANT:  SAKAKI, YOSHIYUKI
;  APPLICANT:  HATTORI, MASAHIRA
;  TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
;  FILE REFERENCE: 249-262
;  CURRENT APPLICATION NUMBER: US/10/156,761
;  CURRENT FILING DATE:  2002-05-29
;  PRIOR APPLICATION NUMBER: JP 2001-204089
;  PRIOR FILING DATE: 2001-05-30
;  PRIOR APPLICATION NUMBER: JP 2001-272697
;  PRIOR FILING DATE: 2001-08-02
;  NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 11984
;   LENGTH: 157
;   TYPE: PRT
;   ORGANISM: Streptomyces avermitilis
US-10-156-761-11984

  Query Match              3.3%;  Score 7;  DB 14;  Length 157;
  Best Local Similarity   100.0%;  Pred. No. 1.4e+02;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        100 RLVSEDP 106
              |||||||
Db         84 RLVSEDP 90


RESULT 11
US-09-809-545A-4
; Sequence 4, Application US/09809545A
; Patent No. US20020110804A1
; GENERAL INFORMATION:
;  APPLICANT: Stanton, Lawrence W.
;  APPLICANT:  White, R. Tyler
;  TITLE OF INVENTION: SECRETED FACTORS
;  FILE REFERENCE: SCIOS.017A
;  CURRENT APPLICATION NUMBER: US/09/809,545A
;  CURRENT FILING DATE:  2001-03-14
;  NUMBER OF SEQ ID NOS: 84
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 4
;   LENGTH: 193
;   TYPE: PRT
;   ORGANISM: Rattus norvegicus
US-09-809-545A-4

  Query Match              3.3%;  Score 7;  DB 9;  Length 193;
  Best Local Similarity   100.0%;  Pred. No. 1.7e+02;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        164 GGRFSSG 170
              |||||||
Db        157 GGRFSSG 163


RESULT 12
US-09-764-870-537
; Sequence 537, Application US/09764870
; Patent No. US20020042386A1
; GENERAL INFORMATION:
;  APPLICANT: Rosen et al.
;  TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
;  FILE REFERENCE: PTZ14
;  CURRENT APPLICATION NUMBER: US/09/764,870
;  CURRENT FILING DATE:  2001-01-17
;  Prior application data removed - consult PALM or file wrapper
;  NUMBER OF SEQ ID NOS: 646
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 537
;   LENGTH: 194
;   TYPE: PRT
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: SITE
;   LOCATION: (160)
;   OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
;   NAME/KEY: SITE
;   LOCATION: (165)
;   OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
;   NAME/KEY: SITE
;   LOCATION: (178)
;   OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
;   NAME/KEY: SITE
;   LOCATION: (179)
;   OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
;   NAME/KEY: SITE
;   LOCATION: (194)
;   OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-764-870-537

  Query Match              3.3%;  Score 7;  DB 9;  Length 194;
  Best Local Similarity   100.0%;  Pred. No. 1.7e+02;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         77 DFISLGL 83
              |||||||
Db         83 DFISLGL 89


RESULT 13
US-10-125-540-537
; Sequence 537, Application US/10125540
; Publication No. US20030059875A1
; GENERAL INFORMATION:
;  APPLICANT: Rosen et al.
;  TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
;  FILE REFERENCE: PTZ14C1
;  CURRENT APPLICATION NUMBER: US/10/125,540
;  CURRENT FILING DATE:  2002-04-19
;  Prior Application removed - See File Wrapper or Palm
;  NUMBER OF SEQ ID NOS: 646
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 537
;   LENGTH: 194
;   TYPE: PRT
```

```
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (160)
;   OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (165)
;   OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (178)
;   OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (179)
;   OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (194)
;   OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-10-125-540-537

  Query Match             3.3%;  Score 7;  DB 14;  Length 194;
  Best Local Similarity   100.0%;  Pred. No. 1.7e+02;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         77 DFISLGL 83
              |||||||
Db         83 DFISLGL 89


RESULT 14
US-09-764-870-432
; Sequence 432, Application US/09764870
; Patent No. US20020042386A1
; GENERAL INFORMATION:
;  APPLICANT: Rosen et al.
;  TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
;  FILE REFERENCE: PTZ14
;  CURRENT APPLICATION NUMBER: US/09/764,870
;  CURRENT FILING DATE:  2001-01-17
;  Prior application data removed - consult PALM or file wrapper
;  NUMBER OF SEQ ID NOS: 646
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 432
;   LENGTH: 210
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-764-870-432

  Query Match             3.3%;  Score 7;  DB 9;  Length 210;
  Best Local Similarity   100.0%;  Pred. No. 1.8e+02;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         77 DFISLGL 83
              |||||||
Db         83 DFISLGL 89


RESULT 15
US-10-125-540-432
; Sequence 432, Application US/10125540
; Publication No. US20030059875A1
; GENERAL INFORMATION:
;  APPLICANT: Rosen et al.
;  TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
;  FILE REFERENCE: PTZ14C1
;  CURRENT APPLICATION NUMBER: US/10/125,540
;  CURRENT FILING DATE:  2002-04-19
;  Prior Application removed - See File Wrapper or Palm
;  NUMBER OF SEQ ID NOS: 646
;  SOFTWARE: PatentIn Ver. 2.0
```

```
; SEQ ID NO 432
;   LENGTH: 210
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-125-540-432

  Query Match             3.3%;  Score 7;  DB 14;  Length 210;
  Best Local Similarity   100.0%;  Pred. No. 1.8e+02;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         77 DFISLGL 83
              |||||||
Db         83 DFISLGL 89

Search completed: March  9, 2004, 17:34:06
Job time : 15.4128 secs
```

```
                         GenCore version 5.1.6
               Copyright (c) 1993 - 2004  Compugen Ltd.

OM protein - protein search, using sw model

Run on:          March  9, 2004, 17:23:39 ; Search time 7.18098 Seconds
                                            (without alignments)
                                            2813.016 Million cell updates/sec

Title:           US-10-006-011A-10
Perfect score:   210
Sequence:        1 GIAESDWHLEGSGGNDAPGQ.........QPLDLQHRAQAGANTRPCPS 210

Scoring table:   OLIGO
                 Gapop 60.0 , Gapext 60.0

Searched:        283366 seqs, 96191526 residues

Word size :        0

Total number of hits satisfying chosen parameters:      283366

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :        PIR_78:*
                 1:  pir1:*
                 2:  pir2:*
                 3:  pir3:*
                 4:  pir4:*

      Pred. No. is the number of results predicted by chance to have a
      score greater than or equal to the score of the result being printed,
      and is derived by analysis of the total score distribution.

                                  SUMMARIES
                    %
Result            Query
  No.    Score    Match Length DB  ID                         Description
 ------------------------------------------------------------------------------
     1      210   100.0   4391  2  A38096                     perlecan precursor
     2       28    13.3   3707  2  S18252                     heparan sulfate pr
     3        8     3.8    497  2  AH2015                     sodium/solute symp
     4        8     3.8    500  2  F70012                     leucyl aminopeptid
     5        8     3.8   1751  1  MMHUMH                     laminin alpha-2 ch
     6        8     3.8   3106  1  S53868                     laminin alpha-2 ch
     7        7     3.3     93  2  S75571                     hypothetical prote
     8        7     3.3    181  2  E87418                     chemotaxis protein
     9        7     3.3    201  2  E75567                     hypothetical prote
    10        7     3.3    210  2  T03144                     hypothetical prote
    11        7     3.3    221  2  T24494                     hypothetical prote
    12        7     3.3    239  2  AE0347                     two-component syst
    13        7     3.3    240  1  JS0591                     endo-1,4-beta-xyla
    14        7     3.3    240  1  S47512                     endo-1,4-beta-xyla
    15        7     3.3    240  2  A82783                     hypothetical prote
    16        7     3.3    241  2  T37005                     endo-1,4-beta-xyla
    17        7     3.3    244  2  T24631                     hypothetical prote
    18        7     3.3    255  2  A35255                     chlorocatechol 1,2
    19        7     3.3    257  2  AB2771                     hypothetical prote
    20        7     3.3    257  2  B97551                     general l-amino ac
    21        7     3.3    273  2  B71355                     probable ribosomal
    22        7     3.3    275  2  AH0335                     probable penicilli
    23        7     3.3    283  2  F83874                     nickel ABC transpo
    24        7     3.3    305  2  B55228                     methionyl-tRNA for
    25        7     3.3    347  2  T35013                     probable membrane
    26        7     3.3    348  2  B58892                     NADH2 dehydrogenas
    27        7     3.3    373  2  F72602                     probable acetylpol
    28        7     3.3    399  2  T01387                     oxidoreductase hom
    29        7     3.3    400  2  D75331                     hypothetical prote
    30        7     3.3    413  2  T29505                     hypothetical prote
    31        7     3.3    434  2  F75425                     tRNA nucleotidyltr
    32        7     3.3    434  2  T31151                     replication protei
    33        7     3.3    445  2  S73596                     nonspecified amino
    34        7     3.3    447  2  C64243                     aminopeptidase (EC
    35        7     3.3    454  2  F90602                     aminopeptidase (le
    36        7     3.3    455  2  H82881                     cytosol aminopepti
    37        7     3.3    468  2  F81800                     leucyl aminopeptid
    38        7     3.3    475  2  T01260                     probable ammonium
    39        7     3.3    479  2  B87699                     conserved hypothet
    40        7     3.3    483  2  T48328                     importin alpha-lik
    41        7     3.3    491  1  C64137                     leucyl aminopeptid
    42        7     3.3    491  2  F82843                     aminopeptidase A/I
    43        7     3.3    499  2  G84972                     leucyl aminopeptid
    44        7     3.3    500  2  A40631                     leucyl aminopeptid
    45        7     3.3    500  2  H97722                     leucyl aminopeptid

                                  ALIGNMENTS

RESULT 1
A38096
perlecan precursor - human
N;Alternate names: basement membrane heparan sulfate proteoglycan; heparan sulfate prote
C;Species: Homo sapiens (man)
C;Date: 07-Apr-1994 #sequence_revision 07-Apr-1994 #text_change 05-Nov-1999
C;Accession: A38096; S19256; S77946; A41059; A40306; B33625; A33625; A41736
R;Murdoch, A.D.; Dodge, G.R.; Cohen, I.; Tuan, R.S.; Iozzo, R.V.
J. Biol. Chem. 267, 8544-8557, 1992
A;Title: Primary structure of the human heparan sulfate proteoglycan from basement membr
tor, laminin, neural cell adhesion molecules, and epidermal growth factor.
A;Reference number: A38096; MUID:92235084; PMID:1569102
A;Accession: A38096
A;Molecule type: mRNA
A;Residues: 1-4391 <MUR>
A;Cross-references: GB:M85289; NID:g184426; PIDN:AAA52700.1; PID:g184427
R;Kallunki, P.; Tryggvason, K.
J. Cell Biol. 116, 559-571, 1992
A;Title: Human basement membrane heparan sulfate proteoglycan core protein: a 467-kD pro
ell adhesion molecules, and epidermal growth factor.
A;Reference number: A41736; MUID:92112994; PMID:1730768
A;Accession: S19256
A;Molecule type: mRNA
A;Residues: 1-57,'D',59-434,'A',436,'FL',438-449,'Q',451-502,'A',503-792,'K',794-908,'R',
71-2979,'H',2981-2994,'G',2996-3167,'T',3169-3240,'R',3242-3426,'R',3428-3631,'Q',3633-3
A;Cross-references: EMBL:X62515
R;Tryggvason, K.
submitted to the EMBL Data Library, October 1991
A;Reference number: S77946
A;Accession: S77946
A;Molecule type: mRNA
A;Residues: 1-57,'D',59-434,'A',436,'FL',438-449,'Q',451-502,'A',503-792,'K',794-908,'R',
71-2979,'H',2981-2994,'G',2996-3167,'T',3169-3240,'R',3242-3426,'R',3428-3631,'Q',3633-4
A;Cross-references: EMBL:X62515; NID:g29469; PIDN:CAA44373.1; PID:g29470
R;Kallunki, P.; Eddy, R.L.; Byers, M.G.; Kestilae, M.; Shows, T.B.; Tryggvason, K.
Genomics 11, 389-396, 1991
A;Title: Cloning of human heparan sulfate proteoglycan core protein, assignment of the g
A;Reference number: A41059; MUID:92120660; PMID:1685141
A;Accession: A41059
A;Molecule type: mRNA
A;Residues: 'RT',892-908,'R',910-1101,'L',1103-1132,'L',1134-1221,'L',1223-1397 <KA2>
A;Cross-references: GB:S76436; NID:g243370; PIDN:AAB21121.1; PID:g243371
R;Dodge, G.R.; Kovalszky, I.; Chu, M.L.; Hassell, J.R.; McBride, O.W.; Yi, H.F.; Iozzo,
Genomics 10, 673-680, 1991
A;Title: Heparan sulfate proteoglycan of human colon: partial molecular cloning, cellul
A;Reference number: A40306; MUID:91365376; PMID:1679749
A;Accession: A40306
A;Molecule type: mRNA
A;Residues: 1018-1405,'G',1407-1409,'G',1411-1465 <DOD>
A;Cross-references: GB:M64283; NID:g184424; PIDN:AAA52699.1; PID:g184425
R;Heremans, A.; van der Schueren, B.; De Cock, B.; Paulsson, M.; Cassiman, J.J.; van den
J. Cell Biol. 109, 3199-3211, 1989
```

```
A;Title: Matrix-associated heparan sulfate proteoglycan: core protein-specific monoclona
anes.
A;Reference number: A33625; MUID:90078352; PMID:2687294
A;Accession: B33625
A;Molecule type: protein
A;Residues: 1379-1384,'X',1386-1388,'X',1390-1398 <HE2>
A;Accession: A33625
A;Molecule type: protein
A;Residues: 2166-2171,'X',2173-2175,'X',2177-2185 <HE3>
A;Note: peptide potentially matches four different regions of sequence shown
C;Genetics:
A;Gene: GDB:HSPG2
A;Cross-references: GDB:126372; OMIM:142461
A;Map position: 1p36.1-1p36.1
C;Superfamily: LDL receptor ligand-binding repeat homology; EGF homology; laminin G repe
C;Keywords: chondroitin sulfate proteoglycan; glycoprotein; heparan sulfate; transmembra
F;1-21/Domain: signal sequence #status predicted <SIG>
F;22-4391/Product: perlecan #status predicted <MAT>
F;22-193/Domain: I <DOM1>
F;194-530/Domain: II <DOM2>
F;199-234/Domain: LDL receptor ligand-binding repeat homology <LDL1>
F;285-319/Domain: LDL receptor ligand-binding repeat homology <LDL2>
F;325-359/Domain: LDL receptor ligand-binding repeat homology <LDL3>
F;368-403/Domain: LDL receptor ligand-binding repeat homology <LDL4>
F;531-1676/Domain: III <DOM3>
F;1159-1206/Domain: laminin-type EGF-like homology <LEG>
F;1563-1610/Domain: laminin-type EGF-like homology <EG7>
F;1613-1668/Domain: laminin-type EGF-like homology <LEG8>
F;1677-3686/Domain: IV <DOM4>
F;2007-2034/Domain: transmembrane #status predicted <TRM>
F;3687-4391/Domain: V <DOM5>
F;3845-3880/Domain: EGF homology <EGF1>
F;3888-3921/Domain: EGF homology <EGF>
F;3953-4106/Domain: laminin G repeat homology <LG2>
F;4147-4175/Domain: EGF homology <EGF2>
F;4149-4151/Region: motor neuron attachment (L-R-E) motif
F;4299-4301/Region: motor neuron attachment (L-R-E) motif
F;65,71,76/Binding site: heparan sulfate (Ser) (covalent) #status predicted
F;89,554,1755,2121,3072,3105,3279,3780,3836,4068/Binding site: carbohydrate (Asn) (coval
F;2995,3933,4179/Binding site: chondroitin sulfate (Ser) (covalent) #status predicted

  Query Match             100.0%;  Score 210;  DB 2;  Length 4391;
  Best Local Similarity   100.0%;  Pred. No. 1.1e-212;
  Matches  210;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 GIAESDWHLEGSGGNDAPGQYGAYFHDDGFLAFPGHVFSRSLPEVPETIELEVRTSTASG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4182 GIAESDWHLEGSGGNDAPGQYGAYFHDDGFLAFPGHVFSRSLPEVPETIELEVRTSTASG 4241

Qy         61 LLLWQGVEVGEAGQGKDFISLGLQDGHLVFRYQLGSGEARLVSEDPINDGEWHRVTALRE 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4242 LLLWQGVEVGEAGQGKDFISLGLQDGHLVFRYQLGSGEARLVSEDPINDGEWHRVTALRE 4301

Qy        121 GRRGSIQVDGEELVSGRSPGPNVAVNAKGSVYIGGAPDVATLTGGRFSSGITGCVKNLVL 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4302 GRRGSIQVDGEELVSGRSPGPNVAVNAKGSVYIGGAPDVATLTGGRFSSGITGCVKNLVL 4361

Qy        181 HSARPGAPPPQPLDLQHRAQAGANTRPCPS 210
              ||||||||||||||||||||||||||||||
Db       4362 HSARPGAPPPQPLDLQHRAQAGANTRPCPS 4391


RESULT 2
S18252
heparan sulfate proteoglycan - mouse
N;Alternate names: perlecan
C;Species: Mus musculus (house mouse)
C;Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 05-Nov-1999
C;Accession: S18252; A31917; B31917; S66460
R;Noonan, D.M.; Fulle, A.; Valente, P.; Cai, S.; Horigan, E.; Sasaki, M.; Yamada, Y.; Ha
J. Biol. Chem. 266, 22939-22947, 1991
A;Title: The complete sequence of perlecan, a basement membrane heparan sulfate proteogl
adhesion molecule.
```

```
A;Reference number: S18252; MUID:92078153; PMID:1744087
A;Accession: S18252
A;Molecule type: mRNA
A;Residues: 1-3707 <NOO>
A;Cross-references: EMBL:M77174; NID:g200295; PIDN:AAA39911.1; PID:g200296
R;Noonan, D.M.; Horigan, E.A.; Ledbetter, S.R.; Vogeli, G.; Sasaki, M.; Yamada, Y.; Hass
J. Biol. Chem. 263, 16379-16387, 1988
A;Title: Identification of cDNA clones encoding different domains of the basement membra
A;Reference number: A92680; MUID:89034110; PMID:2972708
A;Accession: A31917
A;Molecule type: mRNA
A;Residues: 940-1601 <NO2>
A;Cross-references: GB:J04054; NID:g200252; PIDN:AAA39899.1; PID:g200253
A;Accession: B31917
A;Molecule type: mRNA
A;Residues: 1870-2600 <NO3>
A;Cross-references: GB:J04055; NID:g200300; PIDN:AAA39912.1; PID:g200301
R;Schulze, B.; Mann, K.; Battistutta, R.; Wiedemann, H.; Timpl, R.
Eur. J. Biochem. 231, 551-556, 1995
A;Title: Structural properties of recombinant domain III-3 of perlecan containing a glob
A;Reference number: S66460; MUID:95377282; PMID:7649154
A;Accession: S66460
A;Molecule type: protein
A;Residues: 1272-1274,'X',1276,'X',1278-1279 <SCH>
C;Superfamily: LDL receptor ligand-binding repeat homology; EGF homology; laminin G repe
C;Keywords: glycoprotein
F;199-234/Domain: LDL receptor ligand-binding repeat homology <LDL1>
F;285-319/Domain: LDL receptor ligand-binding repeat homology <LDL2>
F;325-359/Domain: LDL receptor ligand-binding repeat homology <LDL3>
F;368-403/Domain: LDL receptor ligand-binding repeat homology <LDL4>
F;764-811/Domain: laminin-type EGF-like homology <LEG>
F;1159-1206/Domain: laminin-type EGF-like homology <LEG7>
F;1563-1610/Domain: laminin-type EGF-like homology <EG7>
F;1613-1668/Domain: laminin-type EGF-like homology <LEG8>
F;3163-3198/Domain: EGF homology <EGF>
F;3270-3423/Domain: laminin G repeat homology <LG2>
F;3464-3492/Domain: EGF homology <EGF7>
F;1256,1891,2336,2394,2427/Binding site: carbohydrate (Asn) (covalent) #status predicted

  Query Match             13.3%;  Score 28;  DB 2;  Length 3707;
  Best Local Similarity   100.0%;  Pred. No. 2.5e-20;
  Matches   28;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        183 ARPGAPPPQPLDLQHRAQAGANTRPCPS 210
              ||||||||||||||||||||||||||||
Db       3680 ARPGAPPPQPLDLQHRAQAGANTRPCPS 3707


RESULT 3
AH2015
sodium/solute symporter [imported] - Nostoc sp. (strain PCC 7120)
C;Species: Nostoc sp. PCC 7120
A;Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
C;Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 09-Dec-2002
C;Accession: AH2015
R;Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi
 Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S
DNA Res. 8, 205-213, 2001
A;Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Ana
A;Reference number: AB1807; MUID:21595285; PMID:11759840
A;Accession: AH2015
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-497 <KUR>
A;Cross-references: GB:BA000019; PIDN:BAB78044.1; PID:g17135498; GSPDB:GN00179
A;Experimental source: strain PCC 7120
C;Genetics:
A;Gene: all1678

  Query Match             3.8%;  Score 8;  DB 2;  Length 497;
  Best Local Similarity   100.0%;  Pred. No. 5.6;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;
```

```
Qy          56 STASGLLL 63
                  ||||||||
Db         325 STASGLLL 332


RESULT 4
F70012
leucyl aminopeptidase homolog yuiE - Bacillus subtilis
C;Species: Bacillus subtilis
C;Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 03-Mar-2003
C;Accession: F70012
R;Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Berte
C.; Bron, S.; Brouillet, S.; Bruschi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Ch
A.; Ehrlich, S.D.; Emmerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari, E.
Nature 390, 249-256, 1997
A;Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Galle
iech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M.F.
Koetter, P.; Koningstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois,
A;Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Maueel
y, M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetelle
 Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadaie, Y.; Sato, T.; Scanlon,
A;Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Sero
akeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiyama,
T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida,
A;Authors: Yoshikawa, H.F.; Zumstein, E.; Yoshikawa, H.; Danchin, A.
A;Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis.
A;Reference number: A69580; MUID:98044033; PMID:9384377
A;Accession: F70012
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-500 <KUN>
A;Cross-references: GB:Z99120; GB:AL009126; NID:g2635613; PIDN:CAB15195.1; PID:g2635702
A;Experimental source: strain 168
C;Genetics:
A;Gene: yuiE
C;Superfamily: Cytosol aminopeptidase

  Query Match             3.8%;  Score 8;  DB 2;  Length 500;
  Best Local Similarity   100.0%;  Pred. No. 5.6;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         158 DVATLTGG 165
                  ||||||||
Db         367 DVATLTGG 374


RESULT 5
MMHUMH
laminin alpha-2 chain - human (fragment)
N;Alternate names: laminin M chain; merosin heavy chain
C;Species: Homo sapiens (man)
C;Date: 30-Sep-1991 #sequence_revision 21-Aug-1998 #text_change 10-Dec-1999
C;Accession: PX0082; A35899; A38970; S14461
R;Hori, H.; Kanamori, T.; Mizuta, T.; Yamaguchi, N.; Liu, Y.; Nagai, Y.
J. Biochem. 116, 1212-1219, 1994
A;Title: Human laminin M chain: Epitope analysis of its monoclonal antibodies by immuno
A;Reference number: PX0082; MUID:95221315; PMID:7535762
A;Accession: PX0082
A;Molecule type: mRNA
A;Residues: 1-1751 <HOR>
A;Experimental source: placenta
R;Ehrig, K.; Leivo, I.; Argraves, W.S.; Ruoslahti, E.; Engvall, E.
Proc. Natl. Acad. Sci. U.S.A. 87, 3264-3268, 1990
A;Title: Merosin, a tissue-specific basement membrane protein, is a laminin-like protei
A;Reference number: A35899; MUID:90238994; PMID:2185464
A;Accession: A35899
A;Status: nucleic acid sequence not shown
A;Molecule type: mRNA
A;Residues: 'V',623-1751 <EHR1>
A;Cross-references: EMBL:M59832
A;Accession: A38970
A;Molecule type: protein
A;Residues: 1368-1384;1389-1406;1593-1607 <EHR2>
```

```
A;Note: the sequence from Fig. 1 is inconsistent with that from Fig. 2 in lacking 1599-V
R;Ehrig, K.; Leivo, I.; Argraves, S.W.; Ruoslahti, E.; Engvall, E.
submitted to the EMBL Data Library, December 1990
A;Description: The tissue-specific basement membrane protein merosin is a laminin-like p
A;Reference number: S14461
A;Accession: S14461
A;Molecule type: mRNA
A;Residues: 'V',623-1264,'R',1266-1751 <LEI>
A;Cross-references: EMBL:M59832; NID:g187520; PIDN:AAA63215.1; PID:g187521
C;Comment: This protein is a prominent component of the basement membrane that mediates
C;Genetics:
A;Gene: GDB:LAMA2; LAMM
A;Cross-references: GDB:132362; OMIM:156225
A;Map position: 6q22-6q23
C;Complex: Laminins are trimers of an alpha-type, a beta-type, and a gamma-type laminin
C;Function:
A;Description: interact with cells and with other basement membrane proteins to promote
C;Superfamily: laminin alpha-1 chain; laminin G repeat homology; laminin-type EGF-like h
C;Keywords: basement membrane; calcium binding; coiled coil; extracellular matrix; glyco
F;21-58/Domain: laminin-type EGF-like homology #status atypical <LE01>
F;61-107/Domain: laminin-type EGF-like homology <LE02>
F;110-165/Domain: laminin-type EGF-like homology <LE03>
F;168-212/Domain: laminin-type EGF-like homology <LE04>
F;527-567,1071-1300/Region: 3DM and 2D9 binding
F;811-972/Domain: laminin G repeat homology <LG1>
F;1005-1165/Domain: laminin G repeat homology <LG2>
F;1191-1354/Domain: laminin G repeat homology <LG3>
F;1430-1578/Domain: laminin G repeat homology <LG4>
F;1605-1751/Domain: laminin G repeat homology <LG5>
F;120,238,255,341,451,542,557,561,658,669,686,767,881,1001,1076,1119,1192,1199,1289,1509,

  Query Match             3.8%;  Score 8;  DB 1;  Length 1751;
  Best Local Similarity   100.0%;  Pred. No. 18;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          48 TIELEVRT 55
                  ||||||||
Db        1429 TIELEVRT 1436


RESULT 6
S53868
laminin alpha-2 chain precursor - mouse
N;Alternate names: laminin M chain; merosin heavy chain
C;Species: Mus musculus (house mouse)
C;Date: 21-Aug-1998 #sequence_revision 21-Aug-1998 #text_change 10-Dec-1999
C;Accession: I49077; S50829; I48655; S31576; S53868
R;Bernier, S.M.; Utani, A.; Sugiyama, S.; Doi, T.; Polistina, C.; Yamada, Y.
Matrix Biol. 14, 447-455, 1995
A;Title: Cloning and expression of laminin alpha 2 chain (M-chain) in the mouse.
A;Reference number: I49077; MUID:95316259; PMID:7795883
A;Accession: I49077
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-3106 <RES>
A;Cross-references: EMBL:U12147; NID:g699109; PIDN:AAC52165.1; PID:g699110
R;Xu, H!; Wu, X.R.; Wewer, U.M.; Engvall, E.
Nature Genet. 8, 297-302, 1994
A;Title: Murine muscular dystrophy caused by a mutation in the laminin alpha-2 (Lama2) g
A;Reference number: S50829; MUID:95179178; PMID:7874173
A;Accession: S50829
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 64-281 <XUH>
A;Cross-references: GB:S75315; NID:g833929; PIDN:AAB33573.1; PID:g833930
R;Chang, A.C.; Wadsworth, S.; Coligan, J.E.
J. Immunol. 151, 1789-1801, 1993
A;Title: Expression of merosin in the thymus and its interaction with thymocytes.
A;Reference number: I48655; MUID:93346725; PMID:8345183
A;Accession: I48655
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 2162-2204,'D',2206-2213,'EY',2216-2279 <RE2>
```

```
A;Cross-references: EMBL:X69869; NID:g53055; PIDN:CAA49502.1; PID:g53056
C;Complex: Laminins are trimers of an alpha-type, a beta-type, and a gamma-type laminin
C;Function:
A;Description: interact with cells and with other basement membrane proteins to promote
C;Superfamily: laminin alpha-1 chain; laminin G repeat homology; laminin-type EGF-like h
C;Keywords: basement membrane; calcium binding; coiled coil; extracellular matrix; glyco
F;1-22/Domain: signal sequence #status predicted <SIG>
F;23-3106/Product: laminin alpha-2 chain #status predicted <MAT>
F;283-337/Domain: laminin-type EGF-like homology <LE01>
F;340-407/Domain: laminin-type EGF-like homology <LE02>
F;410-462/Domain: laminin-type EGF-like homology <LE03>
F;465-511/Domain: laminin-type EGF-like homology <LE04>
F;514-523/Domain: laminin-type EGF-like homology #status atypical <LE05>
F;720-750/Domain: laminin-type EGF-like homology <LE06>
F;753-800/Domain: laminin-type EGF-like homology <LE07>
F;803-858/Domain: laminin-type EGF-like homology <LE08>
F;861-911/Domain: laminin-type EGF-like homology <LE09>
F;914-960/Domain: laminin-type EGF-like homology <LE10>
F;963-1007/Domain: laminin-type EGF-like homology <LE11>
F;1010-1053/Domain: laminin-type EGF-like homology <LE12>
F;1056-1099/Domain: laminin-type EGF-like homology <LE13>
F;1102-1121/Domain: laminin-type EGF-like homology #status atypical <LE14>
F;1123-1159/Domain: laminin-type EGF-like homology #status atypical <LE15>
F;1162-1171/Domain: laminin-type EGF-like homology #status atypical <LE16>
F;1376-1413/Domain: laminin-type EGF-like homology #status atypical <LE17>
F;1416-1462/Domain: laminin-type EGF-like homology <LE18>
F;1465-1520/Domain: laminin-type EGF-like homology <LE19>
F;1523-1567/Domain: laminin-type EGF-like homology <LE20>
F;2166-2327/Domain: laminin G repeat homology <LG1>
F;2360-2520/Domain: laminin G repeat homology <LG2>
F;2546-2709/Domain: laminin G repeat homology <LG3>
F;2785-2933/Domain: laminin G repeat homology <LG4>
F;2960-3106/Domain: laminin G repeat homology <LG5>

  Query Match             3.8%;  Score 8;  DB 1;  Length 3106;
  Best Local Similarity   100.0%;  Pred. No. 30;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         48 TIELEVRT 55
              ||||||||
Db       2784 TIELEVRT 2791


RESULT 7
S75571
hypothetical protein ssr1391 - Synechocystis sp. (strain PCC 6803)
C;Species: Synechocystis sp.
A;Variety: PCC 6803
C;Date: 25-Apr-1997 #sequence_revision 25-Apr-1997 #text_change 08-Oct-1999
C;Accession: S75571
R;Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, N.;
o, K.; Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda
DNA Res. 3, 109-136, 1996
A;Title: Sequence analysis of the genome of the unicellular cyanobacterium Synechocystis
s.
A;Reference number: S74322; MUID:97061201; PMID:8905231
A;Accession: S75571
A;Status: nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-93 <KAN>
A;Cross-references: EMBL:D90911; GB:AB001339; NID:g1653083; PIDN:BAA18132.1; PID:d101886
A;Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996
C;Genetics:
A;Start codon: GTG

  Query Match             3.3%;  Score 7;  DB 2;  Length 93;
  Best Local Similarity   100.0%;  Pred. No. 14;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         48 TIELEVR 54
              |||||||
Db         35 TIELEVR 41
```

```
RESULT 8
E87418
chemotaxis protein CheYIII [imported] - Caulobacter crescentus
C;Species: Caulobacter crescentus
C;Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 20-Apr-2001
C;Accession: E87418
R;Nierman, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg, J.
 B.; Laub, M.T.; DeBoy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Kolon
n, J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C.M.
Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001
A;Title: Complete Genome Sequence of Caulobacter crescentus.
A;Reference number: A87249; MUID:21173698; PMID:11259647
A;Accession: E87418
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-181 <STO>
A;Cross-references: GB:AE005673; NID:g13422715; PIDN:AAK23345.1; GSPDB:GN00148
C;Genetics:
A;Gene: CC1364

  Query Match             3.3%;  Score 7;  DB 2;  Length 181;
  Best Local Similarity   100.0%;  Pred. No. 25;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         53 VRTSTAS 59
              |||||||
Db         72 VRTSTAS 78


RESULT 9
E75567
hypothetical protein - Deinococcus radiodurans (strain R1)
C;Species: Deinococcus radiodurans
C;Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 31-Mar-2000
C;Accession: E75567
R;White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.;
, M.; Shen, M.; Vamathevan, J.J.; Lam, P.; McDonald, L.; Utterback, T.; Zalewski, C.; Ma
S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.
Science 286, 1571-1577, 1999
A;Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.
A;Reference number: A75250; MUID:20036896; PMID:10567266
A;Accession: E75567
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-201 <WHI>
A;Cross-references: GB:AE001867; GB:AE000513; NID:g6457693; PIDN:AAF09633.1; PID:g645770
A;Experimental source: strain R1
C;Genetics:
A;Gene: DR0037
A;Map position: 1

  Query Match             3.3%;  Score 7;  DB 2;  Length 201;
  Best Local Similarity   100.0%;  Pred. No. 28;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        133 LVSGRSP 139
              |||||||
Db        108 LVSGRSP 114


RESULT 10
T03144
hypothetical protein A6 - alcelaphine herpesvirus 1
C;Species: alcelaphine herpesvirus 1
C;Date: 24-Mar-1999 #sequence_revision 24-Mar-1999 #text_change 28-Jul-2000
C;Accession: T03144
R;Ensser, A.; Pflanz, R.; Fleckenstein, B.
J. Virol. 71, 6517-6525, 1997
A;Title: Primary structure of the alcelaphine herpesvirus 1 genome.
A;Reference number: Z14840; MUID:97404659; PMID:9261371
A;Accession: T03144
A;Status: preliminary; translated from GB/EMBL/DDBJ
```

```
A;Molecule type: DNA
A;Residues: 1-210 <ENS>
A;Cross-references: EMBL:AF005370; NID:g2337967; PIDN:AAC58096.1; PID:g2338012
C;Superfamily: alcelaphine herpesvirus 1 hypothetical protein A6

  Query Match             3.3%;  Score 7;  DB 2;  Length 210;
  Best Local Similarity   100.0%;  Pred. No. 29;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         42 LPEVPET 48
              |||||||
Db         88 LPEVPET 94

RESULT 11
T24494
hypothetical protein T05A10.4 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 18-Feb-2000
C;Accession: T24494
R;Sulston, J.
submitted to the EMBL Data Library, November 1995
A;Reference number: Z19898
A;Accession: T24494
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-221 <WIL>
A;Cross-references: EMBL:Z68108; PIDN:CAA92137.1; GSPDB:GN00028; CESP:T05A10.4
A;Experimental source: clone T05A10
C;Genetics:
A;Gene: CESP:T05A10.4
A;Map position: X
A;Introns: 18/3; 54/1; 106/3; 142/2; 190/3

  Query Match             3.3%;  Score 7;  DB 2;  Length 221;
  Best Local Similarity   100.0%;  Pred. No. 30;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         10 EGSGGND 16
              |||||||
Db         23 EGSGGND 29

RESULT 12
AE0347
two-component system response regulator baeR [imported] - Yersinia pestis (strain CO92)
C;Species: Yersinia pestis
C;Date: 02-Nov-2001 #sequence_revision 02-Nov-2001 #text_change 09-Nov-2001
C;Accession: AE0347
R;Parkhill, J.; Wren, B.W.; Thomson, N.R.; Titball, R.W.; Holden, M.T.G.; Prentice, M.B.
deno-Tarraga, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.; Dougan, G.;
il, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrell,
Nature 413, 523-527, 2001
A;Title: Genome sequence of Yersinia pestis, the causative agent of plague.
A;Reference number: AB0001; MUID:21470413; PMID:11586360
A;Accession: AE0347
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-239 <KUR>
A;Cross-references: GB:AL590842; PIDN:CAC92104.1; PID:g15980821; GSPDB:GN00175
C;Genetics:
A;Gene: baeR
C;Superfamily: ompR protein; response regulator homology

  Query Match             3.3%;  Score 7;  DB 2;  Length 239;
  Best Local Similarity   100.0%;  Pred. No. 33;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         34 PGHVFSR 40
              |||||||
Db        175 PGHVFSR 181
```

```
RESULT 13
JS0591
endo-1,4-beta-xylanase (EC 3.2.1.8) C precursor - Streptomyces lividans
N;Alternate names: xylanase C
C;Species: Streptomyces lividans
C;Date: 14-Jul-1994 #sequence_revision 22-Nov-1996 #text_change 18-Jun-1999
C;Accession: JS0591; PS0240
R;Shareck, F.; Roy, C.; Yaguchi, M.; Morosoli, R.; Kluepfel, D.
Gene 107, 75-82, 1991
A;Title: Sequences of three genes specifying xylanases in Streptomyees lividans.
A;Reference number: JS0589; MUID:92077439; PMID:1743521
A;Accession: JS0591
A;Molecule type: DNA
A;Residues: 1-240 <SHA>
A;Cross-references: GB:M64553; NID:g153530; PIDN:AAA26836.1; PID:g153531
A;Accession: PS0240
A;Molecule type: protein
A;Residues: 50-80 <SH2>
C;Genetics:
A;Gene: xlnC
C;Function:
A;Description: catalyzes the hydrolysis of 1,4-beta-xylosidic bonds in xylans
A;Pathway: xylan degradation
C;Superfamily: endo-1,4-beta-xylanase; endo-1,4-beta-xylanase homology
C;Keywords: extracellular protein; glycosidase; hydrolase; polysaccharide degradation
F;1-49/Domain: signal sequence #status predicted <SIG>
F;50-240/Product: endo-1,4-beta-xylanase C #status experimental <MAT>
F;62-239/Domain: endo-1,4-beta-xylanase homology <XYL>
F;134,226/Active site: Glu #status predicted

  Query Match             3.3%;  Score 7;  DB 1;  Length 240;
  Best Local Similarity   100.0%;  Pred. No. 33;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         57 TASGLLL 63
              |||||||
Db         37 TASGLLL 43

RESULT 14
S47512
endo-1,4-beta-xylanase (EC 3.2.1.8) precursor - Streptomyces sp.
N;Alternate names: xylanase
C;Species: Streptomyces sp.
C;Date: 13-Jan-1995 #sequence_revision 22-Nov-1996 #text_change 18-Jun-1999
C;Accession: S47512
R;Mazy-Servais, C.; Moreau, A.; Gerard, C.; Dusart, J.
submitted to the EMBL Data Library, August 1994
A;Description: Cloning and sequencing of a xylanase-encoding gene from Streptomyces sp.
A;Reference number: S47512
A;Accession: S47512
A;Molecule type: DNA
A;Residues: 1-240 <MAZ>
A;Cross-references: EMBL:X81045; NID:g531767; PIDN:CAA56935.1; PID:g531768
A;Experimental source: strain EC3
C;Function:
A;Description: catalyzes the hydrolysis of 1,4-beta-xylosidic bonds in xylans
A;Pathway: xylan degradation
C;Superfamily: endo-1,4-beta-xylanase; endo-1,4-beta-xylanase homology
C;Keywords: extracellular protein; glycosidase; hydrolase; polysaccharide degradation
F;1-47/Domain: signal sequence #status predicted <SIG>
F;48-240/Product: endo-1,4-beta-xylanase #status predicted <MAT>
F;62-239/Domain: endo-1,4-beta-xylanase homology <XYL>
F;135,226/Active site: Glu #status predicted

  Query Match             3.3%;  Score 7;  DB 1;  Length 240;
  Best Local Similarity   100.0%;  Pred. No. 33;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         57 TASGLLL 63
              |||||||
Db         37 TASGLLL 43
```

```
RESULT 15
A82783
hypothetical protein XF0635 [imported] - Xylella fastidiosa (strain 9a5c)
C;Species: Xylella fastidiosa
C;Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 20-Aug-2000
C;Accession: A82783
R;anonymous, The Xylella fastidiosa Consortium of the Organization for Nucleotide Sequen
Nature 406, 151-157, 2000
A;Title: The genome sequence of the plant pathogen Xylella fastidiosa.
A;Reference number: A82515; MUID:20365717; PMID:10910347
A;Note: for a complete list of authors see reference number A59328 below
A;Accession: A82783
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-240 <SIM>
A;Cross-references: GB:AE003908; GB:AE003849; NID:g9105496; PIDN:AAF83445.1; GSPDB:GN001
A;Experimental source: strain 9a5c
R;Simpson, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.; A
Briones, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carrer, H
as-Neto, E.; Docena, C.; El-Dorry, H.; Facincani, A.P.; Ferreira, A.J.S.
submitted to GenBank, June 2000
A;Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; Frohm
J.D.; Junqueira, M.L.; Kemper, E.L.; Kitajima, J.P.; Krieger, J.E.; Kuramae, E.E.; Laigr
chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins, E
A;Authors: Martins, E.M.F.; Matsukuma, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C.Y.;
, F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri, D.A
Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sawasak
A;Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silveir
M.; Tsuhako, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.; Z
A;Reference number: A59328
A;Contents: annotation
C;Genetics:
A;Gene: XF0635

  Query Match             3.3%;  Score 7;  DB 2;  Length 240;
  Best Local Similarity   100.0%;  Pred. No. 33;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        114 RVTALRE 120
              |||||||
Db        219 RVTALRE 225

Search completed: March  9, 2004, 17:31:27
Job time : 7.18098 secs
```

```
                        GenCore version 5.1.6
               Copyright (c) 1993 - 2004  Compugen Ltd.

OM protein - protein search, using sw model

Run on:         March  9, 2004, 17:21:48 ; Search time 5.07923 Seconds
                                           (without alignments)
                                           2152.832 Million cell updates/sec

Title:          US-10-006-011A-10
Perfect score:  210
Sequence:       1 GIAESDWHLEGSGGNDAPGQ..........QPLDLQHRAQAGANTRPCPS 210

Scoring table:  OLIGO
                Gapop 60.0 , Gapext 60.0

Searched:       141681 seqs, 52070155 residues

Word size :       0

Total number of hits satisfying chosen parameters:      141681

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :       SwissProt_42:*

     Pred. No. is the number of results predicted by chance to have a
     score greater than or equal to the score of the result being printed,
     and is derived by analysis of the total score distribution.

                              SUMMARIES

                  %
Result          Query
  No.   Score   Match Length DB  ID                     Description
-------------------------------------------------------------------------
    1     210   100.0   4391  1  PGBM_HUMAN             P98160 homo sapien
    2      28    13.3   3707  1  PGBM_MOUSE             Q05793 mus musculu
    3       8     3.8    500  1  AMPA_BACSU             O32106 bacillus su
    4       8     3.8   3106  1  LMA2_MOUSE             Q60675 mus musculu
    5       8     3.8   3110  1  LMA2_HUMAN             P24043 homo sapien
    6       7     3.3    184  1  RS7_THEAC              Q9hly1 thermoplasm
    7       7     3.3    184  1  RS7_THEVO              Q97cd9 thermoplasm
    8       7     3.3    240  1  XYNC_STRLI             P26220 streptomyce
    9       7     3.3    244  1  YRT2_CAEEL             Q10045 caenorhabdi
   10       7     3.3    248  1  SCTM_HUMAN             Q8wvn6 homo sapien
   11       7     3.3    255  1  TFDC_ALCEU             P05403 alcaligenes
   12       7     3.3    273  1  RL2_TREPA              O83222 treponema p
   13       7     3.3    296  1  CYCG_RHOSH             Q53143 rhodobacter
   14       7     3.3    305  1  FMT_THETH              P43523 thermus the
   15       7     3.3    325  1  NAS2_ORYSA             Q9feg8 oryza sativ
   16       7     3.3    332  1  NAS1_ORYSA             Q9sxq7 oryza sativ
   17       7     3.3    348  1  NU2M_LATCH             O03166 latimeria c
   18       7     3.3    429  1  DAD1_RALSO             Q8y0w7 ralstonia s
   19       7     3.3    445  1  AMPA_MYCPN             P75206 mycoplasma
   20       7     3.3    447  1  AMPA_MYCGE             P47631 mycoplasma
   21       7     3.3    468  1  AMPA_NEIMA             Q9jti8 neisseria m
   22       7     3.3    475  1  AMT2_ARATH             Q9m6n7 arabidopsis
   23       7     3.3    490  1  AMPA_XANAC             Q8pgr0 xanthomonas
   24       7     3.3    491  1  AMPA_HAEIN             P45334 haemophilus
   25       7     3.3    491  1  AMPA_XYLFA             Q9ph08 xylella fas
   26       7     3.3    491  1  AMPA_XYLFT             Q87f32 xylella fas
   27       7     3.3    493  1  AMPA_CLOPE             Q8xhi3 clostridium
   28       7     3.3    493  1  AMPA_XANCP             Q8pcr4 xanthomonas
   29       7     3.3    494  1  AMPA_PASMU             P57823 pasteurella
   30       7     3.3    499  1  AMPA_BUCAI             P57448 buchnera ap
   31       7     3.3    500  1  AMA1_SHEON             Q8ei85 shewanella
   32       7     3.3    500  1  AMPA_RICCN             Q92j85 rickettsia
   33       7     3.3    500  1  AMPA_RICPR             P27888 rickettsia
```

```
   34       7     3.3    502  1  AMPA_VIBPA             Q87lg8 vibrio para
   35       7     3.3    502  1  AMPA_VIBVU             Q8dce5 vibrio vuln
   36       7     3.3    503  1  AMPA_ECOLI             P11648 escherichia
   37       7     3.3    503  1  AMPA_SALTI             Q8z116 salmonella
   38       7     3.3    503  1  AMPA_SALTY             Q8zk29 salmonella
   39       7     3.3    503  1  AMPA_VIBCH             Q9k2w5 vibrio chol
   40       7     3.3    506  1  AMPA_RALSO             Q8xwq8 ralstonia s
   41       7     3.3    517  1  AMPA_STRCO             Q9s2q7 streptomyce
   42       7     3.3    525  1  NCAP_RINDR             Q03332 rinderpest
   43       7     3.3    525  1  NCAP_RINDU             P41359 rinderpest
   44       7     3.3    550  1  WRT2_CAEEL             Q94130 caenorhabdi
   45       7     3.3    603  1  CHLE_MOUSE             Q03311 mus musculu


                              ALIGNMENTS


RESULT 1
PGBM_HUMAN
ID   PGBM_HUMAN     STANDARD;      PRT;  4391 AA.
AC   P98160; Q16287; Q9H3V5;
DT   01-OCT-1996 (Rel. 34, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Basement membrane-specific heparan sulfate proteoglycan core
DE   protein precursor (HSPG) (Perlecan) (PLC).
GN   HSPG2.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=92112994; PubMed=1730768;
RA   Kallunki P., Tryggvason K.;
RT   "Human basement membrane heparan sulfate proteoglycan core protein: a
RT   467-kD protein containing multiple domains resembling elements of the
RT   low density lipoprotein receptor, laminin, neural cell adhesion
RT   molecules, and epidermal growth factor.";
RL   J. Cell Biol. 116:559-571(1992).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Colon, and Skin;
RX   MEDLINE=92235084; PubMed=1569102;
RA   Murdoch A.D., Dodge G.R., Cohen I., Tuan R.S., Iozzo R.V.;
RT   "Primary structure of the human heparan sulfate proteoglycan from
RT   basement membrane (HSPG2/perlecan). A chimeric molecule with multiple
RT   domains homologous to the low density lipoprotein receptor, laminin,
RT   neural cell adhesion molecules, and epidermal growth factor.";
RL   J. Biol. Chem. 267:8544-8557(1992).
RN   [3]
RP   SEQUENCE OF 22-4391 FROM N.A., AND VARIANT SJS1 TYR-1532.
RX   MEDLINE=20553141; PubMed=11101850;
RA   Nicole S., Davoine C.-S., Topaloglu H., Cattolico L., Barral D.,
RA   Beighton P., Ben-Hamida C., Hammouda H., Cruaud C., White P.S.,
RA   Samson D., Urtizberea J.A., Lehmann-Horn F., Weissenbach J.,
RA   Hentati F., Fontaine B.;
RT   "Perlecan, the major proteoglycan of basement membranes, is altered in
RT   patients with Schwartz-Jampel syndrome (chondrodystrophic myotonia).";
RL   Nat. Genet. 26:480-483(2000).
RN   [4]
RP   SEQUENCE OF 1016-1470 FROM N.A.
RC   TISSUE=Colon;
RX   MEDLINE=91365376; PubMed=1679749;
RA   Dodge G.R., Kovalszky I., Chu M.L., Hassell J.R., McBride O.W.,
RA   Yi H.F., Iozzo R.V.;
RT   "Heparan sulfate proteoglycan of human colon: partial molecular
RT   cloning, cellular expression, and mapping of the gene (HSPG2) to the
RT   short arm of human chromosome 1.";
RL   Genomics 10:673-680(1991).
RN   [5]
RP   SEQUENCE OF 890-1396 FROM N.A.
RC   TISSUE=Fibrosarcoma;
```

```
RX   MEDLINE=92120660; PubMed=1685141;
RA   Kallunki P., Eddy R.L., Byers M.G., Kestila M., Shows T.B.,
RA   Tryggvason K.;
RT   "Cloning of human heparan sulfate proteoglycan core protein,
RT   assignment of the gene (HSPG2) to 1p36.1-->p35 and identification of
RT   a BamHI restriction fragment length polymorphism.";
RL   Genomics 11:389-396(1991).
RN   [6]
RP   SEQUENCE OF 1-21 FROM N.A.
RX   MEDLINE=94052171; PubMed=8234307;
RA   Cohen I.R., Graessel S., Murdoch A.D., Iozzo R.V.;
RT   "Structural characterization of the complete human perlecan gene and
RT   its promoter.";
RL   Proc. Natl. Acad. Sci. U.S.A. 90:10404-10408(1993).
RN   [7]
RP   CARBOHYDRATE-LINKAGE SITE ASN-2121.
RX   MEDLINE=22660472; PubMed=12754519;
RA   Zhang H., Li X.-J., Martin D.B., Aebersold R.;
RT   "Identification and quantification of N-linked glycoproteins using
RT   hydrazide chemistry, stable isotope labeling and mass spectrometry.";
RL   Nat. Biotechnol. 21:660-666(2003).
CC   -!- FUNCTION: This protein is an integral component of basement
CC       membranes. It is responsible for the fixed negative electrostatic
CC       charge and is involved in the charge-selective ultrafiltration
CC       properties. It serves as an attachment substrate for cells.
CC   -!- SUBUNIT: Purified perlecan has a strong tendency to aggregate in
CC       dimers or stellate structures. It interacts with other basement
CC       membrane components such as laminin, prolargin and collagen type
CC       IV.
CC   -!- SUBCELLULAR LOCATION: Extracellular.
CC   -!- TISSUE SPECIFICITY: Found in the basement membranes.
CC   -!- PTM: CONTAINS THREE HEPARAN SULFATE CHAINS AS WELL AS N-LINKED
CC       AND O-LINKED OLIGOSACCHARIDES.
CC   -!- DISEASE: Defects in HSPG2 are the cause of Schwartz-Jampel
CC       syndrome (SJS1) [MIM:255800]; a rare autosomal recessive disorder
CC       characterized by permanent myotonia (prolonged failure of muscle
CC       relaxation) and skeletal dysplasia, resulting in reduced stature,
CC       kyphoscoliosis, bowing of the diaphyses and irregular epiphyses.
CC   -!- SIMILARITY: Contains 4 LDL-receptor class A domains.
CC   -!- SIMILARITY: Contains 11 laminin EGF-like domains.
CC   -!- SIMILARITY: Contains 3 laminin IV domains.
CC   -!- SIMILARITY: Contains 22 immunoglobulin-like C2-type domains.
CC   -!- SIMILARITY: Contains 3 laminin G-like domains.
CC   -!- SIMILARITY: Contains 4 EGF-like domains.
CC   -!- SIMILARITY: Contains 1 SEA domain.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; X62515; CAA44373.1; -.
DR   EMBL; M85289; AAA52700.1; -.
DR   EMBL; AL445795; CAC18534.1; -.
DR   EMBL; M64283; AAA52699.1; -.
DR   EMBL; S76436; AAB21121.2; -.
DR   EMBL; L22078; -; NOT_ANNOTATED_CDS.
DR   PIR; A38096; A38096.
DR   HSSP; P00740; 1EDM.
DR   Siena-2DPAGE; P98160; -.
DR   Genew; HGNC:5273; HSPG2.
DR   MIM; 142461; -.
DR   MIM; 255800; -.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR000742; EGF_2.
DR   InterPro; IPR006209; EGF_like.
DR   InterPro; IPR006210; IEGF.
DR   InterPro; IPR007110; Ig-like.
DR   InterPro; IPR003599; Ig.
DR   InterPro; IPR003598; Ig_c2.
```

```
DR   InterPro; IPR003596; Ig_v.
DR   InterPro; IPR000034; Laminin_B.
DR   InterPro; IPR002049; Laminin_EGF.
DR   InterPro; IPR001791; Laminin_G.
DR   InterPro; IPR002172; LDL_receptor_A.
DR   InterPro; IPR000082; SEA_domain.
DR   Pfam; PF00008; EGF; 4.
DR   Pfam; PF00047; ig; 22.
DR   Pfam; PF00052; laminin_B; 3.
DR   Pfam; PF00053; laminin_EGF; 7.
DR   Pfam; PF00054; laminin_G; 3.
DR   Pfam; PF00057; ldl_recept_a; 4.
DR   Pfam; PF01390; SEA; 1.
DR   PRINTS; PR00261; LDLRECEPTOR.
DR   ProDom; PD003031; Laminin_B; 3.
DR   SMART; SM00181; EGF; 15.
DR   SMART; SM00180; EGF_Lam; 12.
DR   SMART; SM00409; IG; 22.
DR   SMART; SM00408; IGc2; 21.
DR   SMART; SM00406; IGv; 7.
DR   SMART; SM00281; LamB; 3.
DR   SMART; SM00282; LamG; 3.
DR   SMART; SM00192; LDLa; 4.
DR   SMART; SM00200; SEA; 1.
DR   PROSITE; PS00022; EGF_1; 9.
DR   PROSITE; PS01186; EGF_2; 6.
DR   PROSITE; PS50026; EGF_3; 4.
DR   PROSITE; PS50835; IG_LIKE; 22.
DR   PROSITE; PS50025; LAM_G_DOMAIN; 3.
DR   PROSITE; PS01248; LAMININ_TYPE_EGF; 11.
DR   PROSITE; PS01209; LDLRA_1; 4.
DR   PROSITE; PS50068; LDLRA_2; 4.
DR   PROSITE; PS50024; SEA; 1.
KW   Signal; Basement membrane; Proteoglycan; Repeat; Glycoprotein;
KW   Heparan sulfate; Laminin EGF-like domain; Immunoglobulin domain;
KW   Extracellular matrix; EGF-like domain; Disease mutation.
FT   SIGNAL        1     21       POTENTIAL.
FT   CHAIN        22   4391       BASEMENT MEMBRANE-SPECIFIC HEPARAN
FT                                SULFATE PROTEOGLYCAN CORE PROTEIN.
FT   DOMAIN      80    194       SEA.
FT   DOMAIN     198    235       LDL-RECEPTOR CLASS A 1.
FT   DOMAIN     284    320       LDL-RECEPTOR CLASS A 2.
FT   DOMAIN     324    360       LDL-RECEPTOR CLASS A 3.
FT   DOMAIN     367    404       LDL-RECEPTOR CLASS A 4.
FT   DOMAIN     405    504       IG-LIKE C2-TYPE 1.
FT   DOMAIN     521    530       LAMININ EGF-LIKE 1 (N-TERMINAL).
FT   DOMAIN     531    730       LAMININ DOMAIN IV 1 (DOMAIN III A).
FT   DOMAIN     731    763       LAMININ EGF-LIKE 1 (C-TERMINAL).
FT   DOMAIN     764    813       LAMININ EGF-LIKE 2.
FT   DOMAIN     814    871       LAMININ EGF-LIKE 3.
FT   DOMAIN     879    923       LAMININ EGF-LIKE 4 (INCOMPLETE).
FT   DOMAIN     924    933       LAMININ EGF-LIKE 5 (N-TERMINAL).
FT   DOMAIN     934   1125       LAMININ DOMAIN IV 2 (DOMAIN III B).
FT   DOMAIN    1126   1158       LAMININ EGF-LIKE 5 (C-TERMINAL).
FT   DOMAIN    1159   1208       LAMININ EGF-LIKE 6.
FT   DOMAIN    1209   1265       LAMININ EGF-LIKE 7.
FT   DOMAIN    1275   1324       LAMININ EGF-LIKE 8.
FT   DOMAIN    1325   1334       LAMININ EGF-LIKE 9 (N-TERMINAL).
FT   DOMAIN    1335   1529       LAMININ DOMAIN IV 3 (DOMAIN III C).
FT   DOMAIN    1530   1562       LAMININ EGF-LIKE 9 (C-TERMINAL).
FT   DOMAIN    1563   1612       LAMININ EGF-LIKE 10.
FT   DOMAIN    1613   1670       LAMININ EGF-LIKE 11.
FT   DOMAIN    1677   1771       IG-LIKE C2-TYPE 2.
FT   DOMAIN    1772   1865       IG-LIKE C2-TYPE 3.
FT   DOMAIN    1866   1955       IG-LIKE C2-TYPE 4.
FT   DOMAIN    1956   2051       IG-LIKE C2-TYPE 5.
FT   DOMAIN    2052   2151       IG-LIKE C2-TYPE 6.
FT   DOMAIN    2152   2244       IG-LIKE C2-TYPE 7.
FT   DOMAIN    2245   2340       IG-LIKE C2-TYPE 8.
FT   DOMAIN    2341   2436       IG-LIKE C2-TYPE 9.
FT   DOMAIN    2437   2533       IG-LIKE C2-TYPE 10.
FT   DOMAIN    2534   2629       IG-LIKE C2-TYPE 11.
FT   DOMAIN    2630   2726       IG-LIKE C2-TYPE 12.
```

```
  Query Match             100.0%;  Score 210;  DB 1;  Length 4391;
  Best Local Similarity   100.0%;  Pred. No. 3.4e-207;
  Matches  210;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 GIAESDWHLEGSGGNDAPGQYGAYFHDDGFLAFPGHVFSRSLPEVPETIELEVRTSTASG 60
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4182 GIAESDWHLEGSGGNDAPGQYGAYFHDDGFLAFPGHVFSRSLPEVPETIELEVRTSTASG 4241

Qy         61 LLLWQGVEVGEAGQGKDFISLGLQDGHLVFRYQLGSGEARLVSEDPINDGEWHRVTALRE 120
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4242 LLLWQGVEVGEAGQGKDFISLGLQDGHLVFRYQLGSGEARLVSEDPINDGEWHRVTALRE 4301

Qy        121 GRRGSIQVDGEELVSGRSPGPNVAVNAKGSVYIGGAPDVATLTGGRFSSGITGCVKNLVL 180
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4302 GRRGSIQVDGEELVSGRSPGPNVAVNAKGSVYIGGAPDVATLTGGRFSSGITGCVKNLVL 4361

Qy        181 HSARPGAPPPQPLDLQHRAQAGANTRPCPS 210
                ||||||||||||||||||||||||||||||
Db       4362 HSARPGAPPPQPLDLQHRAQAGANTRPCPS 4391


RESULT 2
PGBM_MOUSE
ID   PGBM_MOUSE     STANDARD;      PRT;  3707 AA.
AC   Q05793;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Basement membrane-specific heparan sulfate proteoglycan core
DE   protein precursor (HSPG) (Perlecan) (PLC).
GN   HSPG2.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Melanoma;
RX   MEDLINE=92078153; PubMed=1744087;
RA   Noonan D.M., Fulle A., Valente P., Cai S., Horigan E., Sasaki M.,
RA   Yamada Y., Hassell J.R.;
RT   "The complete sequence of perlecan, a basement membrane heparan
RT   sulfate proteoglycan, reveals extensive similarity with laminin A
RT   chain, low density lipoprotein-receptor, and the neural cell adhesion
RT   molecule.";
RL   J. Biol. Chem. 266:22939-22947(1991).
RN   [2]
RP   SEQUENCE OF 940-1601 AND 1870-2600 FROM N.A., AND PARTIAL SEQUENCE.
RX   MEDLINE=89034110; PubMed=2972708;
RA   Noonan D.M., Horigan E.A., Ledbetter S.R., Vogeli G., Sasaki M.,
RA   Yamada Y., Hassell J.R.;
RT   "Identification of cDNA clones encoding different domains of the
RT   basement membrane heparan sulfate proteoglycan.";
RL   J. Biol. Chem. 263:16379-16387(1988).
CC   -!- FUNCTION: This protein is an integral component of basement
CC       membranes. It is responsible for the fixed negative electrostatic
CC       charge and is involved in the charge-selective ultrafiltration
CC       properties. It serves as an attachment substrate for cells.
CC   -!- SUBUNIT: Purified perlecan has a strong tendency to aggregate in
CC       dimers or stellate structures. It interacts with other basement
CC       membrane components such as laminin, prolargin and collagen type
CC       IV.
CC   -!- SUBCELLULAR LOCATION: Extracellular.
CC   -!- TISSUE SPECIFICITY: Found in the basement membranes.
CC   -!- PTM: CONTAINS THREE HEPARAN SULFATE CHAINS AS WELL AS N-LINKED
CC       AND O-LINKED OLIGOSACCHARIDES.
CC   -!- SIMILARITY: Contains 4 LDL-receptor class A domains.
CC   -!- SIMILARITY: Contains 11 laminin EGF-like domains.
CC   -!- SIMILARITY: Contains 3 laminin IV domains.
CC   -!- SIMILARITY: Contains 15 immunoglobulin-like C2-type domains.
CC   -!- SIMILARITY: Contains 3 laminin G-like domains.
CC   -!- SIMILARITY: Contains 1 EGF-like domain.
```

```
CC   -!- SIMILARITY: Contains 1 SEA domain.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; M77174; AAA39911.1; -.
DR   EMBL; J04054; AAA39899.1; -.
DR   EMBL; J04055; AAA39912.1; -.
DR   PIR; S18252; S18252.
DR   PDB; 1GL4; 28-NOV-01.
DR   MGD; MGI:96257; Hspg2.
DR   GO; GO:0005604; C:basement membrane; IDA.
DR   GO; GO:0008104; P:protein localization; IMP.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR000742; EGF_2.
DR   InterPro; IPR006209; EGF_like.
DR   InterPro; IPR007110; Ig-like.
DR   InterPro; IPR003598; Ig_c2.
DR   InterPro; IPR000034; Laminin_B.
DR   InterPro; IPR002049; Laminin_EGF.
DR   InterPro; IPR001791; Laminin_G.
DR   InterPro; IPR002172; LDL_receptor_A.
DR   InterPro; IPR000082; SEA_domain.
DR   Pfam; PF00008; EGF; 4.
DR   Pfam; PF00047; ig; 15.
DR   Pfam; PF00052; laminin_B; 3.
DR   Pfam; PF00053; laminin_EGF; 7.
DR   Pfam; PF00054; laminin_G; 3.
DR   Pfam; PF00057; ldl_recept_a; 4.
DR   Pfam; PF01390; SEA; 1.
DR   PRINTS; PR00261; LDLRECEPTOR.
DR   ProDom; PD003031; Laminin_B; 3.
DR   SMART; SM00180; EGF_Lam; 7.
DR   SMART; SM00408; IGc2; 14.
DR   SMART; SM00281; LamB; 3.
DR   SMART; SM00282; LamG; 3.
DR   SMART; SM00192; LDLa; 4.
DR   SMART; SM00200; SEA; 1.
DR   PROSITE; PS00022; EGF_1; 8.
DR   PROSITE; PS01186; EGF_2; 5.
DR   PROSITE; PS50026; EGF_3; 4.
DR   PROSITE; PS50835; IG_LIKE; 15.
DR   PROSITE; PS50025; LAM_G_DOMAIN; 3.
DR   PROSITE; PS01248; LAMININ_TYPE_EGF; 11.
DR   PROSITE; PS01209; LDLRA_1; 4.
DR   PROSITE; PS50068; LDLRA_2; 4.
DR   PROSITE; PS50024; SEA; 1.
KW   Signal; Basement membrane; Proteoglycan; Repeat; Glycoprotein;
KW   Heparan sulfate; Laminin EGF-like domain; Immunoglobulin domain;
KW   Extracellular matrix; EGF-like domain; 3D-structure.
FT   SIGNAL        1     21       POTENTIAL.
FT   CHAIN        22   3707       BASEMENT MEMBRANE-SPECIFIC HEPARAN
FT                                SULFATE PROTEOGLYCAN CORE PROTEIN.
FT   DOMAIN       80    194       SEA.
FT   DOMAIN      195    234       LDL-RECEPTOR CLASS A 1.
FT   DOMAIN      281    319       LDL-RECEPTOR CLASS A 2.
FT   DOMAIN      320    359       LDL-RECEPTOR CLASS A 3.
FT   DOMAIN      360    403       LDL-RECEPTOR CLASS A 4.
FT   DOMAIN      404    504       IG-LIKE C2-TYPE 1.
FT   DOMAIN      521    530       LAMININ EGF-LIKE 1 (N-TERMINAL).
FT   DOMAIN      531    730       LAMININ DOMAIN IV 1 (DOMAIN III A).
FT   DOMAIN      731    763       LAMININ EGF-LIKE 1 (C-TERMINAL).
FT   DOMAIN      764    813       LAMININ EGF-LIKE 2.
FT   DOMAIN      814    871       LAMININ EGF-LIKE 3.
FT   DOMAIN      879    923       LAMININ EGF-LIKE 4 (INCOMPLETE).
FT   DOMAIN      924    933       LAMININ EGF-LIKE 5 (N-TERMINAL).
FT   DOMAIN      934   1125       LAMININ DOMAIN IV 2 (DOMAIN III B).
FT   DOMAIN     1126   1158       LAMININ EGF-LIKE 5 (C-TERMINAL).
```

```
FT   DOMAIN     1159   1208       LAMININ EGF-LIKE 6.
FT   DOMAIN     1209   1265       LAMININ EGF-LIKE 7.
FT   DOMAIN     1275   1324       LAMININ EGF-LIKE 8.
FT   DOMAIN     1325   1334       LAMININ EGF-LIKE 9 (N-TERMINAL).
FT   DOMAIN     1335   1529       LAMININ DOMAIN IV 3 (DOMAIN III C).
FT   DOMAIN     1530   1562       LAMININ EGF-LIKE 9 (C-TERMINAL).
FT   DOMAIN     1563   1612       LAMININ EGF-LIKE 10.
FT   DOMAIN     1613   1670       LAMININ EGF-LIKE 11.
FT   DOMAIN     1677   1771       IG-LIKE C2-TYPE 2.
FT   DOMAIN     1772   1865       IG-LIKE C2-TYPE 3.
FT   DOMAIN     1866   1954       IG-LIKE C2-TYPE 4.
FT   DOMAIN     1955   2049       IG-LIKE C2-TYPE 5.
FT   DOMAIN     2050   2148       IG-LIKE C2-TYPE 6.
FT   DOMAIN     2149   2244       IG-LIKE C2-TYPE 7.
FT   DOMAIN     2245   2343       IG-LIKE C2-TYPE 8.
FT   DOMAIN     2344   2436       IG-LIKE C2-TYPE 9.
FT   DOMAIN     2437   2532       IG-LIKE C2-TYPE 10.
FT   DOMAIN     2533   2619       IG-LIKE C2-TYPE 11.
FT   DOMAIN     2620   2720       IG-LIKE C2-TYPE 12.
FT   DOMAIN     2721   2809       IG-LIKE C2-TYPE 13.
FT   DOMAIN     2810   2895       IG-LIKE C2-TYPE 14.
FT   DOMAIN     2896   2980       IG-LIKE C2-TYPE 15.
FT   DOMAIN     2984   3162       LAMININ G-LIKE 1.
FT   DOMAIN     3163   3241       EGF-LIKE.
FT   DOMAIN     3245   3425       LAMININ G-LIKE 2.
FT   DOMAIN     3518   3705       LAMININ G-LIKE 3.
FT   SITE         65     67       HEPARAN SULFATE (POTENTIAL).
FT   SITE         71     73       HEPARAN SULFATE (POTENTIAL).
FT   SITE         76     78       HEPARAN SULFATE (POTENTIAL).
FT   SITE       3615   3617       MEDIATES MOTOR NEURON ATTACHMENT
FT                                (POTENTIAL).
FT   DISULFID    199    212       BY SIMILARITY.
FT   DISULFID    206    225       BY SIMILARITY.
FT   DISULFID    219    234       BY SIMILARITY.
FT   DISULFID    285    297       BY SIMILARITY.
FT   DISULFID    292    310       BY SIMILARITY.
FT   DISULFID    304    319       BY SIMILARITY.
FT   DISULFID    325    337       BY SIMILARITY.
FT   DISULFID    332    350       BY SIMILARITY.
FT   DISULFID    344    359       BY SIMILARITY.
FT   DISULFID    368    381       BY SIMILARITY.
FT   DISULFID    375    394       BY SIMILARITY.
FT   DISULFID    388    403       BY SIMILARITY.
FT   DISULFID    428    479       BY SIMILARITY.
FT   DISULFID    764    773       BY SIMILARITY.
FT   DISULFID    766    780       BY SIMILARITY.
FT   DISULFID    783    792       BY SIMILARITY.
FT   DISULFID    795    811       BY SIMILARITY.
FT   DISULFID    814    829       BY SIMILARITY.
FT   DISULFID    816    839       BY SIMILARITY.
FT   DISULFID    842    851       BY SIMILARITY.
FT   DISULFID    854    869       BY SIMILARITY.
FT   DISULFID   1159   1168       BY SIMILARITY.
FT   DISULFID   1161   1175       BY SIMILARITY.
FT   DISULFID   1178   1187       BY SIMILARITY.
FT   DISULFID   1190   1206       BY SIMILARITY.
FT   DISULFID   1209   1224       BY SIMILARITY.
FT   DISULFID   1211   1234       BY SIMILARITY.
FT   DISULFID   1237   1246       BY SIMILARITY.
FT   DISULFID   1249   1263       BY SIMILARITY.
FT   DISULFID   1275   1287       BY SIMILARITY.
FT   DISULFID   1277   1293       BY SIMILARITY.
FT   DISULFID   1295   1304       BY SIMILARITY.
FT   DISULFID   1307   1322       BY SIMILARITY.
FT   DISULFID   1563   1572       BY SIMILARITY.
FT   DISULFID   1565   1579       BY SIMILARITY.
FT   DISULFID   1582   1591       BY SIMILARITY.
FT   DISULFID   1594   1610       BY SIMILARITY.
FT   DISULFID   1613   1628       BY SIMILARITY.
FT   DISULFID   1615   1638       BY SIMILARITY.
FT   DISULFID   1641   1650       BY SIMILARITY.
FT   DISULFID   1653   1668       BY SIMILARITY.
FT   DISULFID   1792   1839       BY SIMILARITY.
```

```
FT   DISULFID   1886   1932       BY SIMILARITY.
FT   DISULFID   1976   2021       BY SIMILARITY.
FT   DISULFID   2073   2118       BY SIMILARITY.
FT   DISULFID   2170   2215       BY SIMILARITY.
FT   DISULFID   2268   2313       BY SIMILARITY.

  Query Match             13.3%;  Score 28;  DB 1;  Length 3707;
  Best Local Similarity   100.0%;  Pred. No. 6.6e-20;
  Matches   28;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        183 ARPGAPPPQPLDLQHRAQAGANTRPCPS 210
              ||||||||||||||||||||||||||||
Db       3680 ARPGAPPPQPLDLQHRAQAGANTRPCPS 3707


RESULT 3
AMPA_BACSU
ID   AMPA_BACSU     STANDARD;      PRT;   500 AA.
AC   O32106;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Probable cytosol aminopeptidase (EC 3.4.11.1) (Leucine aminopeptidase)
DE   (LAP) (Leucyl aminopeptidase).
GN   PEPA OR BSU32050.
OS   Bacillus subtilis.
OC   Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX   NCBI_TaxID=1423;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=168;
RX   MEDLINE=98044033; PubMed=9384377;
RA   Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
RA   Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,
RA   Borriss R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,
RA   Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,
RA   Choi S.K., Codani J.J., Connerton I.F., Cummings N.J., Daniel R.A.,
RA   Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emmerson P.T.,
RA   Entian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,
RA   Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,
RA   Ghim S.Y., Glaser P., Goffeau A., Golightly E.J., Grandi G.,
RA   Guiseppi G., Guy B.J., Haga K., Haiech J., Harwood C.R., Henaut A.,
RA   Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,
RA   Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,
RA   Kobayashi Y., Koetter P., Koningstein G., Krogh S., Kumano M.,
RA   Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,
RA   Lee S.M., Levine A., Liu H., Masuda S., Mauel C., Medigue C.,
RA   Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,
RA   Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,
RA   Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,
RA   Presecan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
RA   Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadaie Y.,
RA   Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,
RA   Sekiguchi J., Sekowska A., Seror S.J., Serror P., Shin B.S., Soldo B.,
RA   Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,
RA   Takeuchi M., Tamakoshi A., Tanaka T., Terpstra P., Tognoni A.,
RA   Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,
RA   Viari A., Wambutt R., Wedler E., Wedler H., Weitzenegger T.,
RA   Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
RA   Yoshida K., Yoshikawa H.F., Zumstein E., Yoshikawa H., Danchin A.;
RT   "The complete genome sequence of the Gram-positive bacterium Bacillus
RT   subtilis.";
RL   Nature 390:249-256(1997).
CC   -!- FUNCTION: Presumably involved in the processing and regular
CC       turnover of intracellular proteins. Catalyzes the removal of
CC       unsubstituted N-terminal amino acids from various peptides (By
CC       similarity).
CC   -!- CATALYTIC ACTIVITY: Release of an N-terminal amino acid, Xaa-|-
CC       Xbb-, in which Xaa is preferably Leu, but may be other amino acids
CC       including Pro although not Arg or Lys, and Xbb may be Pro.
CC   -!- COFACTOR: Binds 2 manganese ions per subunit (By similarity).
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC   -!- SIMILARITY: Belongs to peptidase family M17.
```

```
CC   -----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   -----------------------------------------------------------------------
DR   EMBL; Z99120; CAB15195.1; -.
DR   PIR; F70012; F70012.
DR   HSSP; P00727; 1LAM.
DR   MEROPS; M17.UPW; -.
DR   SubtiList; BG13970; pepA.
DR   HAMAP; MF_00181; -; 1.
DR   InterPro; IPR000819; Peptidase_M17_C.
DR   InterPro; IPR008283; Peptidase_M17_N.
DR   Pfam; PF00883; Peptidase_M17; 1.
DR   Pfam; PF02789; Peptidase_M17_N; 1.
DR   PRINTS; PR00481; LAMNOPPTDASE.
DR   PROSITE; PS00631; CYTOSOL_AP; 1.
KW   Hydrolase; Aminopeptidase; Manganese; Complete proteome.
FT   METAL       261    261       MANGANESE 2 (BY SIMILARITY).
FT   METAL       266    266       MANGANESE 1 AND 2 (BY SIMILARITY).
FT   METAL       284    284       MANGANESE 2 (BY SIMILARITY).
FT   METAL       343    343       MANGANESE 1 (BY SIMILARITY).
FT   METAL       345    345       MANGANESE 1 AND 2 (BY SIMILARITY).
FT   ACT_SITE    273    273       POTENTIAL.
FT   ACT_SITE    347    347       POTENTIAL.
SQ   SEQUENCE   500 AA;  53657 MW;  3E82968F66566559 CRC64;

  Query Match             3.8%;  Score 8;  DB 1;  Length 500;
  Best Local Similarity   100.0%;  Pred. No. 4.1;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        158 DVATLTGG 165
              ||||||||
Db        367 DVATLTGG 374


RESULT 4
LMA2_MOUSE
ID   LMA2_MOUSE     STANDARD;      PRT;  3106 AA.
AC   Q60675; Q05003; Q64061;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   15-MAR-2004 (Rel. 43, Last annotation update)
DE   Laminin alpha-2 chain precursor (Laminin M chain) (Merosin heavy
DE   chain).
GN   LAMA2.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=FVB/N; TISSUE=Embryo, and Heart;
RX   MEDLINE=95316259; PubMed=7795883;
RA   Bernier S.M., Utani A., Sugiyama S., Doi T., Polistina C.,
RA   Yamada Y.;
RT   "Cloning and expression of laminin alpha 2 chain (M-chain) in the
RT   mouse.";
RL   Matrix Biol. 14:447-455(1995).
RN   [2]
RP   SEQUENCE OF 2162-2279 FROM N.A.
RC   STRAIN=C57BL/6; TISSUE=Thymus;
RX   MEDLINE=93346725; PubMed=8345183;
RA   Chang A.C., Wadsworth S., Coligan J.E.;
RT   "Expression of merosin in the thymus and its interaction with
RT   thymocytes.";
RL   J. Immunol. 151:1789-1801(1993).
RN   [3]
RP   SEQUENCE OF 64-281 FROM N.A.
RX   MEDLINE=95179178; PubMed=7874173;
RA   Xu H., Wu X.R., Wewer U.M., Engvall E.;
RT   "Murine muscular dystrophy caused by a mutation in the laminin alpha
RT   2 (Lama2) gene.";
RL   Nat. Genet. 8:297-302(1994).
RN   [4]
RP   SEQUENCE OF 20-25.
RX   MEDLINE=21818471; PubMed=11829758;
RA   Garbe J.H., Gohring W., Mann K., Timpl R., Sasaki T.;
RT   "Complete sequence, recombinant analysis and binding to laminins and
RT   sulphated ligands of the N-terminal domains of laminin alpha3B and
RT   alpha5 chains.";
RL   Biochem. J. 362:213-221(2002).
RN   [5]
RP   BINDING TO FBLN1, FBLN2, AND NID2.
RX   MEDLINE=99146904; PubMed=10022829;
RA   Talts J.F., Andac Z., Goehring W., Brancaccio A., Timpl R.;
RT   "Binding of the G domains of laminin alpha1 and alpha2 chains and
RT   perlecan to heparin, sulfatides, alpha-dystroglycan and several
RT   extracellular matrix proteins.";
RL   EMBO J. 18:863-870(1999).
RN   [6]
RP   X-RAY CRYSTALLOGRAPHY (2.35 ANGSTROMS) OF 2932-3106.
RX   MEDLINE=20085745; PubMed=10619025;
RA   Hohenester E., Tisi D., Talts J.F., Timpl R.;
RT   "The crystal structure of a laminin G-like module reveals the
RT   molecular basis of alpha-dystroglycan binding to laminins, perlecan,
RT   and agrin.";
RL   Mol. Cell 4:783-792(1999).
CC   -!- FUNCTION: Binding to cells via a high affinity receptor, laminin
CC       is thought to mediate the attachment, migration and organization
CC       of cells into tissues during embryonic development by interacting
CC       with other extracellular matrix components.
CC   -!- SUBUNIT: Laminin is a complex glycoprotein, consisting of three
CC       different polypeptide chains (alpha, beta, gamma), which are bound
CC       to each other by disulfide bonds into a cross-shaped molecule
CC       comprising one long and three short arms with globules at each
CC       end. The alpha-2 chain is a subunit of laminin-2 (Merosin) and
CC       laminin-4 (S-merosin). Interacts with FBLN1, FBLN2 and NID2.
CC   -!- SUBCELLULAR LOCATION: Extracellular.
CC   -!- TISSUE SPECIFICITY: Found in the basement membranes (major
CC       component).
CC   -!- DOMAIN: The alpha-helical domains I and II are thought to interact
CC       with other laminin chains to form a coiled coil structure.
CC   -!- DOMAIN: Domains VI, IV and G are globular.
CC   -!- DISEASE: Defects in LAMA2 are a cause of murine muscular dystrophy
CC       (dy2J).
CC   -!- SIMILARITY: Contains 1 laminin N-terminal domain.
CC   -!- SIMILARITY: Contains 17 laminin EGF-like domains.
CC   -!- SIMILARITY: Contains 2 laminin IV domains.
CC   -!- SIMILARITY: Contains 5 laminin G-like domains.
CC   -----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   -----------------------------------------------------------------------
DR   EMBL; U12147; AAC52165.1; -.
DR   EMBL; X69869; CAA49502.1; -.
DR   EMBL; S75315; AAB33573.1; -.
DR   PIR; I49077; S53868.
DR   PDB; 1QU0; 03-DEC-99.
DR   PDB; 1DYK; 04-FEB-01.
DR   MGD; MGI:99912; Lama2.
DR   GO; GO:0005604; C:basement membrane; IDA.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR006209; EGF_like.
DR   InterPro; IPR008979; Gal_bind_like.
DR   InterPro; IPR000034; Laminin_B.
DR   InterPro; IPR002049; Laminin_EGF.
```

```
DR   InterPro; IPR001791; Laminin_G.
DR   InterPro; IPR008211; LamNT.
DR   Pfam; PF00052; laminin_B; 2.
DR   Pfam; PF00053; laminin_EGF; 14.
DR   Pfam; PF00054; laminin_G; 5.
DR   Pfam; PF00055; laminin_Nterm; 1.
DR   PRINTS; PR00011; EGFLAMININ.
DR   ProDom; PD003031; Laminin_B; 1.
DR   SMART; SM00180; EGF_Lam; 14.
DR   SMART; SM00281; LamB; 2.
DR   SMART; SM00282; LamG; 5.
DR   SMART; SM00136; LamNT; 1.
DR   PROSITE; PS00022; EGF_1; 11.
DR   PROSITE; PS01186; EGF_2; 3.
DR   PROSITE; PS01248; LAMININ_TYPE_EGF; 14.
DR   PROSITE; PS50025; LAM_G_DOMAIN; 5.
KW   Glycoprotein; Basement membrane; Extracellular matrix; Coiled coil;
KW   Laminin EGF-like domain; Cell adhesion; Repeat; Signal; 3D-structure.
FT   SIGNAL        1     19
FT   CHAIN        20   3106       LAMININ ALPHA-2 CHAIN.
FT   DOMAIN       20    282       LAMININ N-TERMINAL (DOMAIN VI).
FT   DOMAIN      283    339       LAMININ EGF-LIKE 1.
FT   DOMAIN      340    409       LAMININ EGF-LIKE 2.
FT   DOMAIN      410    464       LAMININ EGF-LIKE 3.
FT   DOMAIN      465    513       LAMININ EGF-LIKE 4.
FT   DOMAIN      514    523       LAMININ EGF-LIKE 5 (N-TERMINAL).
FT   DOMAIN      524    719       LAMININ DOMAIN IV 1 (DOMAIN IV B).
FT   DOMAIN      720    752       LAMININ EGF-LIKE 5 (C-TERMINAL).
FT   DOMAIN      753    802       LAMININ EGF-LIKE 6.
FT   DOMAIN      803    860       LAMININ EGF-LIKE 7.
FT   DOMAIN      861    913       LAMININ EGF-LIKE 8.
FT   DOMAIN      914    962       LAMININ EGF-LIKE 9.
FT   DOMAIN      963   1009       LAMININ EGF-LIKE 10.
FT   DOMAIN     1010   1055       LAMININ EGF-LIKE 11.
FT   DOMAIN     1056   1101       LAMININ EGF-LIKE 12.
FT   DOMAIN     1102   1161       LAMININ EGF-LIKE 13.
FT   DOMAIN     1162   1171       LAMININ EGF-LIKE 14 (N-TERMINAL).
FT   DOMAIN     1172   1375       LAMININ DOMAIN IV 2 (DOMAIN IV A).
FT   DOMAIN     1376   1415       LAMININ EGF-LIKE 14 (C-TERMINAL).
FT   DOMAIN     1416   1464       LAMININ EGF-LIKE 15.
FT   DOMAIN     1465   1522       LAMININ EGF-LIKE 16.
FT   DOMAIN     1523   1569       LAMININ EGF-LIKE 17.
FT   DOMAIN     1570   2140       DOMAIN II AND I.
FT   DOMAIN     2141   2324       LAMININ G-LIKE 1.
FT   DOMAIN     2336   2517       LAMININ G-LIKE 2.
FT   DOMAIN     2522   2706       LAMININ G-LIKE 3.
FT   DOMAIN     2759   2930       LAMININ G-LIKE 4.
FT   DOMAIN     2929   3106       LAMININ G-LIKE 5.
FT   DOMAIN     1662   1863       COILED COIL (POTENTIAL).
FT   DOMAIN     1923   2146       COILED COIL (POTENTIAL).
FT   DISULFID    283    292       BY SIMILARITY.
FT   DISULFID    285    303       BY SIMILARITY.
FT   DISULFID    305    314       BY SIMILARITY.
FT   DISULFID    317    337       BY SIMILARITY.
FT   DISULFID    340    349       BY SIMILARITY.
FT   DISULFID    342    374       BY SIMILARITY.
FT   DISULFID    377    386       BY SIMILARITY.
FT   DISULFID    389    407       BY SIMILARITY.
FT   DISULFID    410    422       BY SIMILARITY.
FT   DISULFID    412    438       BY SIMILARITY.
FT   DISULFID    440    449       BY SIMILARITY.
FT   DISULFID    452    462       BY SIMILARITY.
FT   DISULFID    465    478       BY SIMILARITY.
FT   DISULFID    467    482       BY SIMILARITY.
FT   DISULFID    484    493       BY SIMILARITY.
FT   DISULFID    496    511       BY SIMILARITY.
FT   DISULFID    753    762       BY SIMILARITY.
FT   DISULFID    755    769       BY SIMILARITY.
FT   DISULFID    772    781       BY SIMILARITY.
FT   DISULFID    784    800       BY SIMILARITY.
FT   DISULFID    803    818       BY SIMILARITY.
FT   DISULFID    805    828       BY SIMILARITY.
FT   DISULFID    831    840       BY SIMILARITY.
FT   DISULFID    843    858       BY SIMILARITY.
FT   DISULFID    861    875       BY SIMILARITY.
FT   DISULFID    863    882       BY SIMILARITY.
FT   DISULFID    885    894       BY SIMILARITY.
FT   DISULFID    897    911       BY SIMILARITY.
FT   DISULFID    914    926       BY SIMILARITY.
FT   DISULFID    916    933       BY SIMILARITY.
FT   DISULFID    935    944       BY SIMILARITY.
FT   DISULFID    947    960       BY SIMILARITY.
FT   DISULFID    963    975       BY SIMILARITY.
FT   DISULFID    965    981       BY SIMILARITY.
FT   DISULFID    983    992       BY SIMILARITY.
FT   DISULFID    995   1007       BY SIMILARITY.
FT   DISULFID   1010   1019       BY SIMILARITY.
FT   DISULFID   1012   1026       BY SIMILARITY.
FT   DISULFID   1028   1037       BY SIMILARITY.
FT   DISULFID   1040   1053       BY SIMILARITY.
FT   DISULFID   1056   1068       BY SIMILARITY.
FT   DISULFID   1058   1075       BY SIMILARITY.
FT   DISULFID   1077   1086       BY SIMILARITY.
FT   DISULFID   1089   1099       BY SIMILARITY.
FT   DISULFID   1416   1425       BY SIMILARITY.
FT   DISULFID   1418   1432       BY SIMILARITY.

  Query Match             3.8%;  Score 8;  DB 1;  Length 3106;
  Best Local Similarity   100.0%;  Pred. No. 22;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          48 TIELEVRT 55
               ||||||||
Db        2784 TIELEVRT 2791


RESULT 5
LMA2_HUMAN
ID   LMA2_HUMAN     STANDARD;      PRT;  3110 AA.
AC   P24043; Q14736; Q93022;
DT   01-MAR-1992 (Rel. 21, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   15-MAR-2004 (Rel. 43, Last annotation update)
DE   Laminin alpha-2 chain precursor (Laminin M chain) (Merosin heavy
DE   chain).
GN   LAMA2 OR LAMM.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Placenta;
RX   MEDLINE=94124633; PubMed=8294519;
RA   Vuolteenaho R., Nissinen M., Sainio K., Byers M., Eddy R.,
RA   Hirvonen H., Shows T.B., Sariola H., Engvall E., Tryggvason K.;
RT   "Human laminin M chain (merosin): complete primary structure,
RT   chromosomal assignment, and expression of the M and A chain in human
RT   fetal tissues.";
RL   J. Cell Biol. 124:381-394(1994).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=97066955; PubMed=8910357;
RA   Zhang X., Vuolteenaho R., Tryggvason K.;
RT   "Structure of the human laminin alpha2-chain gene (LAMA2), which is
RT   affected in congenital muscular dystrophy.";
RL   J. Biol. Chem. 271:27664-27669(1996).
RN   [3]
RP   SEQUENCE OF 1981-3110 FROM N.A., AND PARTIAL SEQUENCE.
RC   TISSUE=Placenta;
RX   MEDLINE=90238994; PubMed=2185464;
RA   Ehrig K., Leivo I., Argraves W.S., Ruoslahti E., Engvall E.;
RT   "Merosin, a tissue-specific basement membrane protein, is a
RT   laminin-like protein.";
RL   Proc. Natl. Acad. Sci. U.S.A. 87:3264-3268(1990).
RN   [4]
```

```
RP   VARIANTS GLN-545; HIS-619; LEU-919; HIS-2586 AND LYS-2614.
RA   Panicker S.G., Mendell J.T., Chen L., Feng B., Sahenk Z.,
RA   Marzluf G.A., Amato A.A., Mendell J.R.;
RT   "Novel single base polymorphisms and rare sequence variants in
RT   the laminin 2-chain coding region detected by RNA/SSCP analysis.";
RL   Hum. Mutat. 13:174-174(1999).
RN   [5]
RP   ERRATUM.
RA   Panicker S.G., Mendell J.T., Chen L., Feng B., Sahenk Z.,
RA   Marzluf G.A., Amato A.A., Mendell J.R.;
RL   Hum. Mutat. 13:340-340(1999).
RN   [6]
RP   VARIANT MDC1A PRO-2564.
RX   MEDLINE=21476011; PubMed=11591858;
RA   He Y., Jones K.J., Vignier N., Morgan G., Chevallay M., Barois A.,
RA   Estournet-Mathiaud B., Hori H., Mizuta T., Tome F.M.S., North K.N.,
RA   Guicheney P.;
RT   "Congenital muscular dystrophy with primary partial laminin alpha-2
RT   chain deficiency: molecular study.";
RL   Neurology 57:1319-1322(2001).
RN   [7]
RP   VARIANTS MDC1A TYR-527 AND ARG-862.
RX   MEDLINE=22439669; PubMed=12552556;
RA   Tezak Z., Prandini P., Boscaro M., Marin A., Devaney J., Marino M.,
RA   Fanin M., Trevisan C.P., Park J., Tyson W., Finkel R., Garcia C.,
RA   Angelini C., Hoffman E.P., Pegoraro E.;
RT   "Clinical and molecular study in congenital muscular dystrophy with
RT   partial laminin alpha-2 (LAMA2) deficiency.";
RL   Hum. Mutat. 21:103-111(2003).
CC   -!- FUNCTION: Binding to cells via a high affinity receptor, laminin
CC       is thought to mediate the attachment, migration and organization
CC       of cells into tissues during embryonic development by interacting
CC       with other extracellular matrix components.
CC   -!- SUBUNIT: Laminin is a complex glycoprotein, consisting of three
CC       different polypeptide chains (alpha, beta, gamma), which are bound
CC       to each other by disulfide bonds into a cross-shaped molecule
CC       comprising one long and three short arms with globules at each
CC       end. The alpha-2 chain is a subunit of laminin-2 (merosin) and
CC       laminin-4 (S-merosin).
CC   -!- SUBCELLULAR LOCATION: Extracellular; found in the basement
CC       membranes (major component).
CC   -!- TISSUE SPECIFICITY: Placenta, striated muscle, peripheral nerve,
CC       cardiac muscle, pancreas, lung, spleen, kidney, adrenal gland,
CC       skin, testis, meninges, choroid plexus, and some other regions of
CC       the brain; not in liver, thymus and bone.
CC   -!- DOMAIN: The alpha-helical domains I and II are thought to interact
CC       with other laminin chains to form a coiled coil structure.
CC   -!- DOMAIN: Domains VI, IV and G are globular.
CC   -!- DISEASE: Defects in LAMA2 are the cause of merosin-deficient
CC       congenital muscular dystrophy type 1A (MDC1A) [MIM:607855]. MDC1A
CC       is characterized by difficulty walking, hypotonia, proximal
CC       weakness, hyporeflexia, and white matter hypodensity on MRI.
CC   -!- SIMILARITY: Contains 1 laminin N-terminal domain.
CC   -!- SIMILARITY: Contains 17 laminin EGF-like domains.
CC   -!- SIMILARITY: Contains 2 laminin IV domains.
CC   -!- SIMILARITY: Contains 5 laminin G-like domains.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; Z26653; CAA81394.1; -.
DR   EMBL; U66796; AAB18388.1; -.
DR   EMBL; U66733; AAB18388.1; JOINED.
DR   EMBL; U66734; AAB18388.1; JOINED.
DR   EMBL; U66735; AAB18388.1; JOINED.
DR   EMBL; U66736; AAB18388.1; JOINED.
DR   EMBL; U66737; AAB18388.1; JOINED.
DR   EMBL; U66738; AAB18388.1; JOINED.
```

```
DR   EMBL; U66739; AAB18388.1; JOINED.
DR   EMBL; U66740; AAB18388.1; JOINED.
DR   EMBL; U66741; AAB18388.1; JOINED.
DR   EMBL; U66742; AAB18388.1; JOINED.
DR   EMBL; U66743; AAB18388.1; JOINED.
DR   EMBL; U66745; AAB18388.1; JOINED.
DR   EMBL; U66746; AAB18388.1; JOINED.
DR   EMBL; U66747; AAB18388.1; JOINED.
DR   EMBL; U66748; AAB18388.1; JOINED.
DR   EMBL; U66749; AAB18388.1; JOINED.
DR   EMBL; U66750; AAB18388.1; JOINED.
DR   EMBL; U66751; AAB18388.1; JOINED.
DR   EMBL; U66752; AAB18388.1; JOINED.
DR   EMBL; U66753; AAB18388.1; JOINED.
DR   EMBL; U66754; AAB18388.1; JOINED.
DR   EMBL; U66755; AAB18388.1; JOINED.
DR   EMBL; U66756; AAB18388.1; JOINED.
DR   EMBL; U66757; AAB18388.1; JOINED.
DR   EMBL; U66758; AAB18388.1; JOINED.
DR   EMBL; U66759; AAB18388.1; JOINED.
DR   EMBL; U66760; AAB18388.1; JOINED.
DR   EMBL; U66761; AAB18388.1; JOINED.
DR   EMBL; U66762; AAB18388.1; JOINED.
DR   EMBL; U66763; AAB18388.1; JOINED.
DR   EMBL; U66764; AAB18388.1; JOINED.
DR   EMBL; U66765; AAB18388.1; JOINED.
DR   EMBL; U66766; AAB18388.1; JOINED.
DR   EMBL; U66768; AAB18388.1; JOINED.
DR   EMBL; U66769; AAB18388.1; JOINED.
DR   EMBL; U66770; AAB18388.1; JOINED.
DR   EMBL; U66771; AAB18388.1; JOINED.
DR   EMBL; U66772; AAB18388.1; JOINED.
DR   EMBL; U66773; AAB18388.1; JOINED.
DR   EMBL; U66774; AAB18388.1; JOINED.
DR   EMBL; U66775; AAB18388.1; JOINED.
DR   EMBL; U66776; AAB18388.1; JOINED.
DR   EMBL; U66777; AAB18388.1; JOINED.
DR   EMBL; U66778; AAB18388.1; JOINED.
DR   EMBL; U66779; AAB18388.1; JOINED.
DR   EMBL; U66780; AAB18388.1; JOINED.
DR   EMBL; U66781; AAB18388.1; JOINED.
DR   EMBL; U66782; AAB18388.1; JOINED.
DR   EMBL; U66783; AAB18388.1; JOINED.
DR   EMBL; U66784; AAB18388.1; JOINED.
DR   EMBL; U66785; AAB18388.1; JOINED.
DR   EMBL; U66786; AAB18388.1; JOINED.
DR   EMBL; U66787; AAB18388.1; JOINED.
DR   EMBL; U66788; AAB18388.1; JOINED.
DR   EMBL; U66789; AAB18388.1; JOINED.
DR   EMBL; U66790; AAB18388.1; JOINED.
DR   EMBL; U66791; AAB18388.1; JOINED.
DR   EMBL; U66792; AAB18388.1; JOINED.
DR   EMBL; U66793; AAB18388.1; JOINED.
DR   EMBL; U66794; AAB18388.1; JOINED.
DR   EMBL; U66795; AAB18388.1; JOINED.
DR   EMBL; M59832; AAA63215.1; -.
DR   PIR; PX0082; MMHUMH.
DR   HSSP; Q60675; 1QU0.
DR   Genew; HGNC:6482; LAMA2.
DR   MIM; 156225; -.
DR   MIM; 607855; -.
DR   GO; GO:0005604; C:basement membrane; TAS.
DR   GO; GO:0005198; F:structural molecule activity; TAS.
DR   GO; GO:0007517; P:muscle development; TAS.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR006209; EGF_like.
DR   InterPro; IPR008979; Gal_bind_like.
DR   InterPro; IPR000034; Laminin_B.
DR   InterPro; IPR002049; Laminin_EGF.
DR   InterPro; IPR001791; Laminin_G.
DR   InterPro; IPR008211; LamNT.
DR   Pfam; PF00052; laminin_B; 2.
DR   Pfam; PF00053; laminin_EGF; 14.
```

```
DR   Pfam; PF00054; laminin_G; 5.
DR   Pfam; PF00055; laminin_Nterm; 1.
DR   PRINTS; PR00011; EGFLAMININ.
DR   ProDom; PD003031; Laminin_B; 1.
DR   SMART; SM00180; EGF_Lam; 15.
DR   SMART; SM00281; LamB; 2.
DR   SMART; SM00282; LamG; 5.
DR   SMART; SM00136; LamNT; 1.
DR   PROSITE; PS00022; EGF_1; 11.
DR   PROSITE; PS01186; EGF_2; 3.
DR   PROSITE; PS01248; LAMININ_TYPE_EGF; 14.
DR   PROSITE; PS50025; LAM_G_DOMAIN; 5.
KW   Glycoprotein; Basement membrane; Extracellular matrix; Coiled coil;
KW   Laminin EGF-like domain; Cell adhesion; Repeat; Signal; Polymorphism;
KW   Disease mutation.
FT   SIGNAL        1     22       POTENTIAL.

  Query Match             3.8%;  Score 8;  DB 1;  Length 3110;
  Best Local Similarity   100.0%;  Pred. No. 22;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         48 TIELEVRT 55
              ||||||||
Db       2788 TIELEVRT 2795


RESULT 6
RS7_THEAC
ID   RS7_THEAC      STANDARD;      PRT;   184 AA.
AC   Q9HLY1;
DT   28-FEB-2003 (Rel. 41, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   30S ribosomal protein S7P.
GN   RPS7P OR TA0092.
OS   Thermoplasma acidophilum.
OC   Archaea; Euryarchaeota; Thermoplasmata; Thermoplasmatales;
OC   Thermoplasmataceae; Thermoplasma.
OX   NCBI_TaxID=2303;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=DSM 1728;
RX   MEDLINE=20479972; PubMed=11029001;
RA   Ruepp A., Graml W., Santos-Martinez M.-L., Koretke K.K., Volker C.,
RA   Mewes H.-W., Frishman D., Stocker S., Lupas A.N., Baumeister W.;
RT   "The genome sequence of the thermoacidophilic scavenger Thermoplasma
RT   acidophilum.";
RL   Nature 407:508-513(2000).
CC   -!- FUNCTION: One of the primary rRNA binding proteins, it binds
CC       directly to 16S rRNA where it nucleates assembly of the head
CC       domain of the 30S subunit. Is located at the subunit interface
CC       close to the decoding center (By similarity).
CC   -!- SUBUNIT: Part of the 30S ribosomal subunit.
CC   -!- SIMILARITY: Belongs to the S7P family of ribosomal proteins.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; AL445063; CAC11240.1; -.
DR   HSSP; P17291; 1RSS.
DR   HAMAP; MF_00480; -; 1.
DR   InterPro; IPR000235; Ribosomal_S7.
DR   InterPro; IPR005716; Ribosomal_S7e/a.
DR   Pfam; PF00177; Ribosomal_S7; 1.
DR   ProDom; PD000817; Ribosomal_S7; 1.
DR   TIGRFAMs; TIGR01028; S7_S5_E_A; 1.
DR   PROSITE; PS00052; RIBOSOMAL_S7; FALSE_NEG.
KW   Ribosomal protein; RNA-binding; rRNA-binding; Complete proteome.
SQ   SEQUENCE   184 AA;  20525 MW;  E64459E7DEE004F1 CRC64;

  Query Match             3.3%;  Score 7;  DB 1;  Length 184;
  Best Local Similarity   100.0%;  Pred. No. 17;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        163 TGGRFSS 169
              |||||||
Db         32 TGGRFSS 38


RESULT 7
RS7_THEVO
ID   RS7_THEVO      STANDARD;      PRT;   184 AA.
AC   Q97CD9;
DT   28-FEB-2003 (Rel. 41, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   30S ribosomal protein S7P.
GN   RPS7P OR TV0163 OR TVG0172869.
OS   Thermoplasma volcanium.
OC   Archaea; Euryarchaeota; Thermoplasmata; Thermoplasmatales;
OC   Thermoplasmataceae; Thermoplasma.
OX   NCBI_TaxID=50339;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=GSS1 / DSM 4299 / JCM 9571;
RX   MEDLINE=20570466; PubMed=11121031;
RA   Kawashima T., Amano N., Koike H., Makino S.-I., Higuchi S.,
RA   Kawashima-Ohya Y., Watanabe K., Yamazaki M., Kanehori K., Kawamoto T.,
RA   Nunoshiba T., Yamamoto Y., Aramaki H., Makino K., Suzuki M.;
RT   "Archaeal adaptation to higher temperatures revealed by genomic
RT   sequence of Thermoplasma volcanium.";
RL   Proc. Natl. Acad. Sci. U.S.A. 97:14257-14262(2000).
CC   -!- FUNCTION: One of the primary rRNA binding proteins, it binds
CC       directly to 16S rRNA where it nucleates assembly of the head
CC       domain of the 30S subunit. Is located at the subunit interface
CC       close to the decoding center (By similarity).
CC   -!- SUBUNIT: Part of the 30S ribosomal subunit.
CC   -!- SIMILARITY: Belongs to the S7P family of ribosomal proteins.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; AP000991; BAB59305.1; -.
DR   HAMAP; MF_00480; -; 1.
DR   InterPro; IPR000235; Ribosomal_S7.
DR   InterPro; IPR005716; Ribosomal_S7e/a.
DR   Pfam; PF00177; Ribosomal_S7; 1.
DR   ProDom; PD000817; Ribosomal_S7; 1.
DR   TIGRFAMs; TIGR01028; S7_S5_E_A; 1.
DR   PROSITE; PS00052; RIBOSOMAL_S7; FALSE_NEG.
KW   Ribosomal protein; RNA-binding; rRNA-binding; Complete proteome.
SQ   SEQUENCE   184 AA;  20474 MW;  BFA9FCE6CE0082AE CRC64;

  Query Match             3.3%;  Score 7;  DB 1;  Length 184;
  Best Local Similarity   100.0%;  Pred. No. 17;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        163 TGGRFSS 169
              |||||||
Db         32 TGGRFSS 38


RESULT 8
XYNC_STRLI
ID   XYNC_STRLI     STANDARD;      PRT;   240 AA.
AC   P26220;
```

```
DT   01-MAY-1992 (Rel. 22, Created)
DT   01-MAY-1992 (Rel. 22, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Endo-1,4-beta-xylanase C precursor (EC 3.2.1.8) (Xylanase C)
DE   (1,4-beta-D-xylan xylanohydrolase C).
GN   XLNC.
OS   Streptomyces lividans.
OC   Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC   Streptomycineae; Streptomycetaceae; Streptomyces.
OX   NCBI_TaxID=1916;
RN   [1]
RP   SEQUENCE FROM N.A., AND SEQUENCE OF 50-80.
RC   STRAIN=66 / 1326;
RX   MEDLINE=92077439; PubMed=1743521;
RA   Shareck F., Roy C., Yaguchi M., Morosoli R., Kluepfel D.;
RT   "Sequences of three genes specifying xylanases in Streptomyces
RT   lividans.";
RL   Gene 107:75-82(1991).
CC   -!- FUNCTION: Contributes to hydrolyze hemicellulose, the major
CC       component of plant cell-walls.
CC   -!- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-xylosidic
CC       linkages in xylans.
CC   -!- PATHWAY: Xylan degradation.
CC   -!- SUBCELLULAR LOCATION: Secreted.
CC   -!- SIMILARITY: Belongs to cellulase family G (family 11 of glycosyl
CC       hydrolases).
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; M64553; AAA26836.1; -.
DR   EMBL; A25307; CAA01768.1; -.
DR   PIR; JS0591; JS0591.
DR   HSSP; P09850; 1XNB.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR001137; Glyco_hydro_11.
DR   InterPro; IPR006311; Tat.
DR   Pfam; PF00457; Glyco_hydro_11; 1.
DR   PRINTS; PR00911; GLHYDRLASE11.
DR   TIGRFAMs; TIGR01409; TAT_signal_seq; 1.
DR   PROSITE; PS00776; GLYCOSYL_HYDROL_F11_1; 1.
DR   PROSITE; PS00777; GLYCOSYL_HYDROL_F11_2; 1.
KW   Xylan degradation; Hydrolase; Glycosidase; Signal.
FT   SIGNAL        1     49
FT   CHAIN        50    240       ENDO-1,4-BETA-XYLANASE C.
FT   ACT_SITE    134    134       NUCLEOPHILE (BY SIMILARITY).
FT   ACT_SITE    226    226       PROTON DONOR (BY SIMILARITY).
SQ   SEQUENCE   240 AA;  25673 MW;  FC663415780142CA CRC64;

  Query Match             3.3%;  Score 7;  DB 1;  Length 240;
  Best Local Similarity   100.0%;  Pred. No. 22;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         57 TASGLLL 63
              |||||||
Db         37 TASGLLL 43

RESULT 9
YRT2_CAEEL
ID   YRT2_CAEEL     STANDARD;      PRT;   244 AA.
AC   Q10045;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   15-JUL-1998 (Rel. 36, Last annotation update)
DE   Hypothetical 28.6 kDa protein T07A5.2 in chromosome III.
GN   T07A5.2.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC   Rhabditidae; Peloderinae; Caenorhabditis.
OX   NCBI_TaxID=6239;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Bristol N2;
RA   Buck D.;
RL   Submitted (FEB-1995) to the EMBL/GenBank/DDBJ databases.
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC   -!- SIMILARITY: TO YEAST YKR030W AND S.POMBE SPBC25H2.14.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; Z48055; CAA88132.1; -.
DR   PIR; T24631; T24631.
DR   WormPep; T07A5.2; CE01647.
DR   InterPro; IPR007881; UNC-50.
DR   Pfam; PF05216; UNC-50; 1.
KW   Hypothetical protein; Transmembrane.
FT   TRANSMEM     59     79       POTENTIAL.
FT   TRANSMEM     86    106       POTENTIAL.
FT   TRANSMEM    137    157       POTENTIAL.
FT   TRANSMEM    165    185       POTENTIAL.
FT   TRANSMEM    198    218       POTENTIAL.
SQ   SEQUENCE   244 AA;  28641 MW;  026DF6A97A7CD613 CRC64;

  Query Match             3.3%;  Score 7;  DB 1;  Length 244;
  Best Local Similarity   100.0%;  Pred. No. 23;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        159 VATLTGG 165
              |||||||
Db        211 VATLTGG 217

RESULT 10
SCTM_HUMAN
ID   SCTM_HUMAN     STANDARD;      PRT;   248 AA.
AC   Q8WVN6; O00466;
DT   28-FEB-2003 (Rel. 41, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Secreted and transmembrane protein 1 precursor (Protein K12).
GN   SECTM1.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A., SUBCELLULAR LOCATION, AND TISSUE SPECIFICITY.
RX   MEDLINE=98149980; PubMed=9480746;
RA   Slentz-Kesler K.A., Hale L.P., Kaufman R.E.;
RT   "Identification and characterization of K12 (SECTM1), a novel human
RT   gene that encodes a Golgi-associated protein with transmembrane and
RT   secreted isoforms.";
RL   Genomics 47:327-340(1998).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Liver;
RX   MEDLINE=22388257; PubMed=12477932;
RA   Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA   Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA   Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA   Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA   Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA   Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA   Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
```

```
RA   Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA   Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA   Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA   Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA   Fahey J., Helton E., Ketteman M., Madan A., Rodrigues S., Sanchez A.,
RA   Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA   Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA   Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA   Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
RA   Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT   "Generation and initial analysis of more than 15,000 full-length
RT   human and mouse cDNA sequences.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN   [3]
RP   INTERACTION WITH CD7.
RX   MEDLINE=20119303; PubMed=10652336;
RA   Lyman S.D., Escobar S., Rousseau A.-M., Armstrong A., Fanslow W.C.;
RT   "Identification of CD7 as a cognate of the human K12 (SECTM1)
RT   protein.";
RL   J. Biol. Chem. 275:3431-3437(2000).
CC   -!- SUBUNIT: Interacts with CD7.
CC   -!- SUBCELLULAR LOCATION: Type I membrane protein (Probable). Also
CC       found as secreted.
CC   -!- TISSUE SPECIFICITY: Detected at the highest levels in peripheral
CC       blood leukocytes and breast cancer cell lines. Found in leukocytes
CC       of the myeloid lineage, with the strongest expression observed in
CC       granulocytes and no detectable expression in lymphocytes.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; U77643; AAC52044.1; -.
DR   EMBL; BC017716; AAH17716.1; -.
DR   Genew; HGNC:10707; SECTM1.
DR   MIM; 602602; -.
DR   InterPro; IPR007110; Ig-like.
KW   Transmembrane; Signal.
FT   SIGNAL        1     28       POTENTIAL.
FT   CHAIN        29    248       SECRETED AND TRANSMEMBRANE PROTEIN 1.
FT   DOMAIN       29    145       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    146    166       POTENTIAL.
FT   DOMAIN      167    248       CYTOPLASMIC (POTENTIAL).
FT   DISULFID     38     55       POTENTIAL.
FT   CARBOHYD     56     56       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CONFLICT    191    191       V -> F (IN REF. 2).
SQ   SEQUENCE   248 AA;  27039 MW;  21E3066B67920487 CRC64;

  Query Match              3.3%;  Score 7;  DB 1;  Length 248;
  Best Local Similarity  100.0%;  Pred. No. 23;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         31 LAFPGHV 37
              |||||||
Db          6 LAFPGHV 12


RESULT 11
TFDC_ALCEU
ID   TFDC_ALCEU     STANDARD;      PRT;   255 AA.
AC   P05403; P71131;
DT   01-NOV-1988 (Rel. 09, Created)
DT   01-AUG-1992 (Rel. 23, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Chlorocatechol 1,2-dioxygenase (EC 1.13.11.1) (Pyrocatechase).
GN   TFDC.
OS   Alcaligenes eutrophus (Ralstonia eutropha), and
OS   Burkholderia cepacia (Pseudomonas cepacia).
OG   Plasmid pJP4, and Plasmid pMAB1.
OC   Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
OC   Burkholderiaceae; Ralstonia.
OX   NCBI_TaxID=510, 292;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   SPECIES=A.eutrophus; STRAIN=JMP134; PLASMID=pJP4;
RX   MEDLINE=90236889; PubMed=2185214;
RA   Perkins E.J., Gordon M.P., Caceres O., Lurquin P.F.;
RT   "Organization and sequence analysis of the 2,4-dichlorophenol
RT   hydroxylase and dichlorocatechol oxidative operons of plasmid pJP4.";
RL   J. Bacteriol. 172:2351-2359(1990).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   SPECIES=A.eutrophus; STRAIN=JMP134; PLASMID=pJP4;
RX   MEDLINE=88303359; PubMed=3405772;
RA   Perkins E.J., Bolton G., Gordon M.P., Lurquin P.F.;
RT   "Partial nucleotide sequence of the chlorocatechol degradative operon
RT   tfdCDEF of pJP4 and similarity to promoters of the chlorinated
RT   aromatic degradative operons tfdA and clcABD.";
RL   Nucleic Acids Res. 16:7200-7200(1988).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   SPECIES=A.eutrophus; PLASMID=pJP4;
RX   MEDLINE=88142559; PubMed=2830460;
RA   Ghosal D., You I.-S.;
RT   "Nucleotide homology and organization of chlorocatechol oxidation
RT   genes of plasmids pJP4 and pAC27.";
RL   Mol. Gen. Genet. 211:113-120(1988).
RN   [4]
RP   SEQUENCE FROM N.A.
RC   SPECIES=B.cepacia; STRAIN=CSV90; PLASMID=pMAB1;
RX   MEDLINE=94161508; PubMed=7509586;
RA   Bhat M.A., Tsuda M., Nozaki M., Horiike K., Vaidyanathan C.S.,
RA   Nakazawa T.;
RT   "Identification and characterization of a new plasmid carrying genes
RT   for degradation of 2,4-dichlorophenoxyacetate from Pseudomonas
RT   cepacia CSV90.";
RL   Appl. Environ. Microbiol. 60:307-312(1994).
CC   -!- FUNCTION: PREFERENTIALLY CONVERTS 3,5-DICHLOROCATECHOL RELATIVE
CC       TO OTHER CHLORINATED CATECHOLS. RETAIN DIMINISHED ACTIVITY TOWARD
CC       NONCHLORINATED SUBSTRATES.
CC   -!- CATALYTIC ACTIVITY: Catechol + O(2) = cis,cis-muconate.
CC   -!- COFACTOR: Binds 1 iron (ferric) ion per subunit.
CC   -!- PATHWAY: 3-chlorocatechol degradation.
CC   -!- SIMILARITY: Belongs to the intradiol ring-cleavage dioxygenase
CC       family.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; M35097; AAA98262.1; -.
DR   EMBL; M36280; AAA98261.1; -.
DR   EMBL; D16356; BAA03859.1; -.
DR   PIR; A35255; A35255.
DR   InterPro; IPR000627; Dioxygenase.
DR   InterPro; IPR007535; Dioxygenase_N.
DR   Pfam; PF00775; Dioxygenase; 1.
DR   Pfam; PF04444; Dioxygenase_N; 1.
DR   PROSITE; PS00083; INTRADIOL_DIOXYGENAS; 1.
KW   Aromatic hydrocarbons catabolism; Oxidoreductase; Dioxygenase; Iron;
KW   Plasmid.
FT   METAL       130    130       IRON (BY SIMILARITY).
FT   METAL       164    164       IRON (BY SIMILARITY).
FT   METAL       188    188       IRON (BY SIMILARITY).
FT   METAL       190    190       IRON (BY SIMILARITY).
FT   CONFLICT    115    116       ED -> DH (IN REF. 2).
SQ   SEQUENCE   255 AA;  28283 MW;  5BCC1BC6A94B19B4 CRC64;
```

```
  Query Match             3.3%;  Score 7;  DB 1;  Length 255;
  Best Local Similarity   100.0%;  Pred. No. 24;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        198 RAQAGAN 204
              |||||||
Db        248 RAQAGAN 254

RESULT 12
RL2_TREPA
ID   RL2_TREPA      STANDARD;      PRT;   273 AA.
AC   O83222;
DT   15-DEC-1998 (Rel. 37, Created)
DT   15-DEC-1998 (Rel. 37, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   50S ribosomal protein L2.
GN   RPLB OR TP0192.
OS   Treponema pallidum.
OC   Bacteria; Spirochaetes; Spirochaetales; Spirochaetaceae; Treponema.
OX   NCBI_TaxID=160;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Nichols;
RX   MEDLINE=98332770; PubMed=9665876;
RA   Fraser C.M., Norris S.J., Weinstock G.M., White O., Sutton G.G.,
RA   Dodson R., Gwinn M., Hickey E.K., Clayton R., Ketchum K.A.,
RA   Sodergren E., Hardham J.M., McLeod M.P., Salzberg S., Peterson J.,
RA   Khalak H., Richardson D., Howell J.K., Chidambaram M., Utterback T.,
RA   McDonald L., Artiach P., Bowman C., Cotton M.D., Fujii C., Garland S.,
RA   Hatch B., Horst K., Roberts K., Sandusky M., Weidman J., Smith H.O.,
RA   Venter J.C.;
RT   "Complete genome sequence of Treponema pallidum, the syphilis
RT   spirochete.";
RL   Science 281:375-388(1998).
CC   -!- FUNCTION: This protein is a primary 23S rRNA-binding protein. It
CC       has peptidyltransferase activity (By similarity).
CC   -!- SIMILARITY: Belongs to the L2P family of ribosomal proteins.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; AE001202; AAC65177.1; -.
DR   PIR; B71355; B71355.
DR   HSSP; P04257; 1RL2.
DR   TIGR; TP0192; -.
DR   InterPro; IPR008994; Nucleic_acid_OB.
DR   InterPro; IPR002171; Ribosomal_L2.
DR   InterPro; IPR005880; Ribosomal_L2_b/o.
DR   InterPro; IPR008991; Transl_SH3_like.
DR   Pfam; PF00181; Ribosomal_L2; 1.
DR   Pfam; PF03947; Ribosomal_L2_C; 1.
DR   TIGRFAMs; TIGR01171; rplB_bact; 1.
DR   PROSITE; PS00467; RIBOSOMAL_L2; FALSE_NEG.
KW   Ribosomal protein; rRNA-binding; Complete proteome.
SQ   SEQUENCE   273 AA;  29867 MW;  66F989616C28B1FE CRC64;

  Query Match             3.3%;  Score 7;  DB 1;  Length 273;
  Best Local Similarity   100.0%;  Pred. No. 25;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         96 SGEARLV 102
              |||||||
Db        178 SGEARLV 184

RESULT 13
CYCG_RHOSH
ID   CYCG_RHOSH     STANDARD;      PRT;   296 AA.
AC   Q53143;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Diheme cytochrome C-type.
GN   CYCG.
OS   Rhodobacter sphaeroides (Rhodopseudomonas sphaeroides).
OC   Bacteria; Proteobacteria; Alphaproteobacteria; Rhodobacterales;
OC   Rhodobacteraceae; Rhodobacter.
OX   NCBI_TaxID=1063;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 17023 / 2.4.1 / NCIB 8253 / DSM 158;
RX   MEDLINE=95362655; PubMed=7543472;
RA   Flory J.E., Donohue T.J.;
RT   "Organization and expression of the Rhodobacter sphaeroides cycFG
RT   operon.";
RL   J. Bacteriol. 177:4311-4320(1995).
CC   -!- FUNCTION: DIHEME C-TYPE CYTOCHROME, THAT IS PARTICULARLY EXPRESSED
CC       WHEN CELLS GENERATE ENERGY VIA AEROBIC RESPIRATION.
CC   -!- SUBCELLULAR LOCATION: Membrane-bound.
CC   -!- PTM: Binds 2 heme groups per molecule (Potential).
CC   -!- SIMILARITY: TO ACETOBACTER ALCOHOL DEHYDROGENASE CYTOCHROME C
CC       SUBUNIT.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; L36880; AAD09146.1; -.
DR   InterPro; IPR003088; Cyt_CI.
DR   InterPro; IPR003219; CytC_adh.
DR   InterPro; IPR000345; CytC_heme_BS.
DR   Pfam; PF00034; cytochrome_c; 1.
DR   ProDom; PD011584; CytC_adh; 1.
DR   PROSITE; PS00190; CYTOCHROME_C; 2.
KW   Electron transport; Heme; Membrane.
FT   BINDING      52     52       HEME 1 (COVALENT) (BY SIMILARITY).
FT   BINDING      55     55       HEME 1 (COVALENT) (BY SIMILARITY).
FT   METAL        56     56       IRON (HEME 1 AXIAL LIGAND)
FT                                (BY SIMILARITY).
FT   BINDING     202    202       HEME 2 (COVALENT) (BY SIMILARITY).
FT   BINDING     205    205       HEME 2 (COVALENT) (BY SIMILARITY).
FT   METAL       206    206       IRON (HEME 2 AXIAL LIGAND)
FT                                (BY SIMILARITY).
SQ   SEQUENCE   296 AA;  31727 MW;  4C4A9D8F695B5BFD CRC64;

  Query Match             3.3%;  Score 7;  DB 1;  Length 296;
  Best Local Similarity   100.0%;  Pred. No. 27;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         96 SGEARLV 102
              |||||||
Db         62 SGEARLV 68

RESULT 14
FMT_THETH
ID   FMT_THETH      STANDARD;      PRT;   305 AA.
AC   P43523;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Methionyl-tRNA formyltransferase (EC 2.1.2.9).
GN   FMT.
OS   Thermus thermophilus.
OC   Bacteria; Deinococcus-Thermus; Deinococci; Thermales; Thermaceae;
OC   Thermus.
```

```
OX   NCBI_TaxID=274;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=VK1;
RX   MEDLINE=95050326; PubMed=7961514;
RA   Meinnel T., Blanquet S.;
RT   "Characterization of the Thermus thermophilus locus encoding peptide
RT   deformylase and methionyl-tRNA(fMet) formyltransferase.";
RL   J. Bacteriol. 176:7387-7390(1994).
CC   -!- FUNCTION: Modify the free amino group of the aminoacyl moiety of
CC       methionyl-tRNA(fMet). The formyl group appears to play a dual role
CC       in the initiator identity of N-formylmethionyl-tRNA by: (I)
CC       promoting its recognition by IF2 and (II) impairing its binding to
CC       EFTu-GTP.
CC   -!- CATALYTIC ACTIVITY: 10-formyltetrahydrofolate + L-methionyl-
CC       tRNA(fMet) + H(2)O = tetrahydrofolate + N-formylmethionyl-
CC       tRNA(fMet).
CC   -!- SIMILARITY: Belongs to the fmt family.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; X79087; CAA55696.1; -.
DR   HSSP; P23882; 1FMT.
DR   HAMAP; MF_00182; -; 1.
DR   InterPro; IPR005794; Fmt.
DR   InterPro; IPR005793; Formyl_trans_C.
DR   InterPro; IPR002376; formyl_transf.
DR   InterPro; IPR001555; GART_AS.
DR   Pfam; PF02911; formyl_trans_C; 1.
DR   Pfam; PF00551; formyl_transf; 1.
DR   TIGRFAMs; TIGR00460; fmt; 1.
DR   PROSITE; PS00373; GART; FALSE_NEG.
KW   Transferase; Methyltransferase; Protein biosynthesis.
FT   BINDING     108    111       TETRAHYDROFOLATE (THF) (BY SIMILARITY).
SQ   SEQUENCE   305 AA;  33323 MW;  1AA26B3FFBC82B46 CRC64;

  Query Match             3.3%;  Score 7;  DB 1;  Length 305;
  Best Local Similarity   100.0%;  Pred. No. 28;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         57 TASGLLL 63
              |||||||
Db        267 TASGLLL 273


RESULT 15
NAS2_ORYSA
ID   NAS2_ORYSA     STANDARD;      PRT;   325 AA.
AC   Q9FEG8;
DT   28-FEB-2003 (Rel. 41, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Probable nicotianamine synthase 2 (EC 2.5.1.43) (S-adenosyl-L-
DE   methionine:S-adenosyl-L-methionine:S-adenosyl-methionine 3-amino-3-
DE   carboxypropyltransferase 2) (OsNAS2).
GN   NAS2.
OS   Oryza sativa (Rice).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC   Ehrhartoideae; Oryzeae; Oryza.
OX   NCBI_TaxID=4530;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=cv. Indica-IR36; TISSUE=Root;
RX   MEDLINE=21097014; PubMed=11169192;
RA   Higuchi K., Watanabe S., Takahashi M., Kawasaki S., Nakanishi H.,
RA   Nishizawa N.-K., Mori S.;
RT   "Nicotianamine synthase gene expression differs in barley and rice
RT   under Fe-deficient conditions.";
RL   Plant J. 25:159-167(2001).
CC   -!- FUNCTION: Synthesizes nicotianamine, a polyamine that is the first
CC       intermediate in the synthesis of the phytosiderophores of the
CC       mugineic acid type found in gramineae which serve as a sensor for
CC       the physiological iron status within the plant, and/or might be
CC       involved in the transport of iron (By similarity).
CC   -!- CATALYTIC ACTIVITY: 3 S-adenosyl-L-methionine = 3 5'-S-methyl-5'-
CC       thioadenosine + nicotianamine.
CC   -!- COFACTOR: Pyridoxal phosphate.
CC   -!- SIMILARITY: Contains 1 NAS domain.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; AB046401; BAB17826.1; -.
DR   EMBL; AB023818; BAB17823.1; -.
DR   Gramene; Q9FEG8; -.
DR   InterPro; IPR004298; Nicotian_synth.
DR   Pfam; PF03059; NAS; 1.
KW   Transferase; Pyridoxal phosphate; Multigene family.
FT   DOMAIN        3    282       NAS.
SQ   SEQUENCE   325 AA;  34454 MW;  04563BFEF5D64BCF CRC64;

  Query Match             3.3%;  Score 7;  DB 1;  Length 325;
  Best Local Similarity   100.0%;  Pred. No. 29;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        158 DVATLTG 164
              |||||||
Db        187 DVATLTG 193


Search completed: March  9, 2004, 17:28:35
Job time : 5.07923 secs
```

```
                    GenCore version 5.1.6
           Copyright (c) 1993 - 2004  Compugen Ltd.

OM protein - protein search, using sw model

Run on:          March  9, 2004, 17:22:43 ; Search time 18.3903 Seconds
                                            (without alignments)
                                            3602.917 Million cell updates/sec

Title:           US-10-006-011A-10
Perfect score:   210
Sequence:        1 GIAESDWHLEGSGGNDAPGQ..........QPLDLQHRAQAGANTRPCPS 210

Scoring table:   OLIGO
                 Gapop 60.0 , Gapext 60.0

Searched:        1017041 seqs, 315518202 residues

Word size :       0

Total number of hits satisfying chosen parameters:      1017041

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :        SPTREMBL_25:*
                1:  sp_archea:*
                2:  sp_bacteria:*
                3:  sp_fungi:*
                4:  sp_human:*
                5:  sp_invertebrate:*
                6:  sp_mammal:*
                7:  sp_mhc:*
                8:  sp_organelle:*
                9:  sp_phage:*
                10:  sp_plant:*
                11:  sp_rodent:*
                12:  sp_virus:*
                13:  sp_vertebrate:*
                14:  sp_unclassified:*
                15:  sp_rvirus:*
                16:  sp_bacteriap:*
                17:  sp_archeap:*

      Pred. No. is the number of results predicted by chance to have a
      score greater than or equal to the score of the result being printed,
      and is derived by analysis of the total score distribution.

                                    SUMMARIES
                       %
Result               Query
  No.    Score   Match Length DB  ID                      Description
 ---------------------------------------------------------------------------
     1      13    6.2     68  11  O08591                   O08591 rattus norv
     2       9    4.3   3367  5   Q9XZC9                  Q9xzc9 drosophila
     3       9    4.3   3375  5   Q8IP51                  Q8ip51 drosophila
     4       8    3.8    494  16   Q81XS5                  Q81xs5 bacillus an
     5       8    3.8    494  16   Q816E3                  Q816e3 bacillus ce
     6       8    3.8    497  2   Q7X2C8                  Q7x2c8 geobacillus
     7       8    3.8    497  16   Q8YWD5                  Q8ywd5 anabaena sp
     8       8    3.8    659  11   Q7TQI9                  Q7tqi9 mus musculu
     9       8    3.8    858  11   Q8R145                  Q8r145 mus musculu
    10       7    3.3     49  8   O79191                  O79191 ninox stren
    11       7    3.3     71  16   Q7UXR0                  Q7uxr0 rhodopirell
    12       7    3.3     91  16   Q82PU0                  Q82pu0 streptomyce
    13       7    3.3     93  16   P74056                  P74056 synechocyst
    14       7    3.3    106  10   Q9XHS4                  Q9xhs4 brassica ca
    15       7    3.3    108  10   Q9XHS6                  Q9xhs6 brassica ol
    16       7    3.3    122  4   Q8N8V6                  Q8n8v6 homo sapien
    17       7    3.3    151  12   Q918M4                  Q918m4 measles vir
    18       7    3.3    152  4   Q8N124                  Q8n124 homo sapien
    19       7    3.3    152  4   Q8N197                  Q8n197 homo sapien
    20       7    3.3    153  5   Q9GPR3                  Q9gpr3 dictyosteli
    21       7    3.3    157  16   Q82F13                  Q82f13 streptomyce
    22       7    3.3    162  16   Q7UDF5                  Q7udf5 shigella fl
    23       7    3.3    165  16   Q7WM03                  Q7wm03 bordetella
    24       7    3.3    172  16   Q7W8E2                  Q7w8e2 bordetella
    25       7    3.3    181  16   Q9A8I9                  Q9a8i9 caulobacter
    26       7    3.3    185  2   Q8KP10                  Q8kp10 bacillus me
    27       7    3.3    186  5   Q8MSM1                  Q8msm1 drosophila
    28       7    3.3    193  11   Q8VD50                  Q8vd50 rattus norv
    29       7    3.3    201  16   Q9RYB3                  Q9ryb3 deinococcus
    30       7    3.3    209  12   Q64922                  Q64922 alcelaphine
    31       7    3.3    209  12   Q64915                  Q64915 alcelaphine
    32       7    3.3    210  12   O36399                  O36399 alcelaphine
    33       7    3.3    217  2   Q9L3H5                  Q9l3h5 uncultured
    34       7    3.3    221  5   P90959                  P90959 caenorhabdi
    35       7    3.3    226  10   Q9ZSV4                  Q9zsv4 acetabulari
    36       7    3.3    233  17   Q8ZVX2                  Q8zvx2 pyrobaculum
    37       7    3.3    239  16   Q8ZCV7                  Q8zcv7 yersinia pe
    38       7    3.3    239  16   Q8CWH6                  Q8cwh6 yersinia pe
    39       7    3.3    240  2   Q56013                  Q56013 streptomyce
    40       7    3.3    240  16   Q9PFM2                  Q9pfm2 xylella fas
    41       7    3.3    241  2   Q7X397                  Q7x397 fritschea b
    42       7    3.3    241  16   Q9RI72                  Q9ri72 streptomyce
    43       7    3.3    242  10   Q84NM0                  Q84nm0 oryza sativ
    44       7    3.3    245  16   Q8NPY7                  Q8npy7 corynebacte
    45       7    3.3    247  16   Q8KEJ3                  Q8kej3 chlorobium


                                 ALIGNMENTS


RESULT 1
O08591
ID   O08591     PRELIMINARY;      PRT;    68 AA.
AC   O08591;
DT   01-JUL-1997 (TrEMBLrel. 04, Created)
DT   01-JUL-1997 (TrEMBLrel. 04, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Perlecan (Fragment).
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Kidney;
RA   Pyke C., Kristensen P., Ostergaard P.B., Oturai P.S., Romer J.;
RT   "Proteoglycan expression in the normal rat kidney.";
RL   Submitted (OCT-1996) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; U75305; AAB51124.1; -.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR001791; Laminin_G.
DR   Pfam; PF00054; laminin_G; 1.
DR   PROSITE; PS50025; LAM_G_DOMAIN; 1.
FT   NON_TER       1      1
FT   NON_TER      68     68
SQ   SEQUENCE   68 AA;  7395 MW;  5868E45D8A7083E0 CRC64;

  Query Match              6.2%;  Score 13;  DB 11;  Length 68;
  Best Local Similarity  100.0%;  Pred. No. 2.1e-05;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         38 FSRSLPEVPETIE 50
              |||||||||||||
Db         21 FSRSLPEVPETIE 33


RESULT 2
Q9XZC9
ID   Q9XZC9      PRELIMINARY;      PRT;  3367 AA.
```

```
AC   Q9XZC9;
DT   01-NOV-1999 (TrEMBLrel. 12, Created)
DT   01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Laminin ALPHA1,2 (SYMBOL=WB).
GN   WB OR WING BLISTER OR CG15288.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC   Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=OREGON-R;
RX   MEDLINE=99207061; PubMed=10189378;
RA   Martin D., Zusman S., Li X., Williams E.L., Khare N., DaRocha S.,
RA   Chiquet-Ehrismann R., Baumgartner S.;
RT   "wing blister, a new Drosophila laminin alpha chain required for cell
RT   adhesion and migration during embryonic and imaginal development.";
RL   J. Cell Biol. 145:191-201(1999).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Berkeley;
RX   MEDLINE=99403001; PubMed=10471707;
RA   Ashburner M., Misra S., Roote J., Lewis S.E., Blazej R., Davis T.,
RA   Doyle C., Galle R., George R., Harris N., Hartzell G., Harvey D.,
RA   Hong L., Houston K., Hoskins R., Johnson G., Martin C., Moshrefi A.,
RA   Palazzolo M., Reese M.G., Spradling A., Tsang G., Wan K., Whitelaw K.,
RA   Celniker S., Rubin G.M.;
RT   "An exploration of the sequence of a 2.9-Mb region of the genome of
RT   Drosophila melanogaster: the Adh region.";
RL   Genetics 153:179-219(1999).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Berkeley;
RA   Celniker S.E., Agbayani A., Arcaina T.T., Baxter E., Blazej R.G.,
RA   Butenhoff C., Champe M., Chavez C., Chew M., Ciesiolka L., Doyle C.M.,
RA   Farfan D.E., Galle R., George R.A., Harris N.L., Hoskins R.A.,
RA   Houston K.A., Hummasti S.R., Karra K., Kearney L., Kim E., Lee B.,
RA   Lewis S., Li P., Lomotan M.A., Mazda P., Moshrefi A.R., Moshrefi M.,
RA   Nixon K., Pacleb J.M., Park S., Pfeiffer B., Poon L., Sequeira A.,
RA   Sethi H., Snir E., Svirskas R.R., Wan K.H., Weinburg T., Zhang R.,
RA   Zieran L.L., Rubin G.M.;
RL   Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF135118; AAD31714.1; -.
DR   EMBL; AE003409; AAF44861.1; -.
DR   HSSP; P00740; 1EDM.
DR   FlyBase; FBgn0004002; wb.
DR   GO; GO:0007267; P:cell-cell signaling; NAS.
DR   InterPro; IPR001589; Actbind_actnin.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR006209; EGF_like.
DR   InterPro; IPR000034; Laminin_B.
DR   InterPro; IPR002049; Laminin_EGF.
DR   InterPro; IPR001791; Laminin_G.
DR   InterPro; IPR008211; LamNT.
DR   Pfam; PF00052; laminin_B; 2.
DR   Pfam; PF00053; laminin_EGF; 14.
DR   Pfam; PF00054; laminin_G; 4.
DR   Pfam; PF00055; laminin_Nterm; 1.
DR   PRINTS; PR00011; EGFLAMININ.
DR   SMART; SM00180; EGF_Lam; 16.
DR   SMART; SM00282; LamG; 5.
DR   SMART; SM00136; LamNT; 1.
DR   PROSITE; PS00019; ACTININ_1; 1.
DR   PROSITE; PS00022; EGF_1; 14.
DR   PROSITE; PS01186; EGF_2; 4.
DR   PROSITE; PS01248; LAMININ_TYPE_EGF; 16.
DR   PROSITE; PS50025; LAM_G_DOMAIN; 4.
KW   Laminin EGF-like domain.
SQ   SEQUENCE   3367 AA;  374097 MW;  EB125654B1BC1511 CRC64;

  Query Match             4.3%;  Score 9;  DB 5;  Length 3367;
  Best Local Similarity   100.0%;  Pred. No. 11;
  Matches    9;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        108 NDGEWHRVT 116
              |||||||||
Db       3081 NDGEWHRVT 3089


RESULT 3
Q8IP51
ID   Q8IP51     PRELIMINARY;      PRT;  3375 AA.
AC   Q8IP51;
DT   01-MAR-2003 (TrEMBLrel. 23, Created)
DT   01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   CG15288-PB.
GN   WB OR CG15288.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC   Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=20196006; PubMed=10731132;
RA   Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA   Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA   George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA   Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA   Brandon R.C., Rogers Y.H., Blazej R.G., Champe M., Pfeiffer B.D.,
RA   Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Gabor G.L.,
RA   Abril J.F., Agbayani A., An H.J., Andrews-Pfannkoch C., Baldwin D.,
RA   Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA   Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA   Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA   Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA   Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA   de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA   Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA   Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA   Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA   Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA   Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA   Hostin D., Houston K.A., Howland T.J., Wei M.H., Ibegwam C.,
RA   Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA   Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA   Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA   Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA   Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA   Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA   Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA   Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA   Reinert K., Remington K., Saunders R.D., Scheeler F., Shen H.,
RA   Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA   Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA   Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA   Wang Z.Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA   Williams S.M., WoodageT, Worley K.C., Wu D., Yang S., Yao Q.A., Ye J.,
RA   Yeh R.F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA   Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA   Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT   "The genome sequence of Drosophila melanogaster.";
RL   Science 287:2185-2195(2000).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Celniker S.E., Adams M.D., Kronmiller B., Wan K.H., Holt R.A.,
RA   Evans C.A., Gocayne J.D., Amanatides P.G., Brandon R.C., Rogers Y.,
RA   Banzon J., An H., Baldwin D., Banzon J., Beeson K.Y., Busam D.A.,
RA   Carlson J.W., Center A., Champe M., Davenport L.B., Dietz S.M.,
RA   Dodson K., Dorsett V., Doup L.E., Doyle C., Dresnek D., Farfan D.,
RA   Ferriera S., Frise E., Galle R.F., Garg N.S., George R.A.,
RA   Gonzalez M., Houck J., Hoskins R.A., Hostin D., Howland T.J.,
RA   Ibegwam C., Jalali M., Kruse D., Li P., Mattei B., Moshrefi A.,
```

```
RA   McIntosh T.C., Moy M., Murphy B., Nelson C., Nelson K.A., Nunoo J.,
RA   Pacleb J., Paragas V., Park S., Patel S., Pfeiffer B.,
RA   Phouanenavong S., Pittman G.S., Puri V., Richards S., Scheeler F.,
RA   Stapleton M., Strong R., Svirskas R., Tector C., Tyler D.,
RA   Williams S.M., Zaveri J.S., Smith H.O., Venter J.C., Rubin G.M.;
RT   "Sequencing of Drosophila melanogaster genome.";
RL   Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RA   Misra S., Crosby M.A., Matthews B.B., Bayraktaroglu L., Campbell K.,
RA   Hradecky P., Huang Y., Kaminker J.S., Prochnik S.E., Smith C.D.,
RA   Tupy J.L., Bergman C., Berman B., Carlson J.W., Celniker S.E.,
RA   Clamp M., Drysdale R., Emmert D., Frise E., de Grey A., Harris N.,
RA   Kronmiller B., Marshall B., Millburn G., Richter J., Russo S.,
RA   Searle S.M.J., Smith E., Shu S., Smutniak F., Whitfield E.,
RA   Ashburner M., Gelbart W.M., Rubin G.M., Mungall C.J., Lewis S.E.;
RT   "Annotation of Drosophila melanogaster genome.";
RL   Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
RN   [4]
RP   SEQUENCE FROM N.A.
RA   Adams M.D., Celniker S.E., Gibbs R.A., Rubin G.M., Venter C.J.;
RL   Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
RN   [5]
RP   SEQUENCE FROM N.A.
RA   FlyBase;
RL   Submitted (SEP-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AE003643; AAN10875.1; -.
DR   FlyBase; FBgn0004002; wb.
DR   GO; GO:0007267; P:cell-cell signaling; NAS.
DR   InterPro; IPR001589; Actbind_actnin.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR006209; EGF_like.
DR   InterPro; IPR006210; IEGF.
DR   InterPro; IPR000034; Laminin_B.
DR   InterPro; IPR002049; Laminin_EGF.
DR   InterPro; IPR001791; Laminin_G.
DR   InterPro; IPR008211; LamNT.
DR   Pfam; PF00052; laminin_B; 2.
DR   Pfam; PF00053; laminin_EGF; 14.
DR   Pfam; PF00054; laminin_G; 4.
DR   Pfam; PF00055; laminin_Nterm; 1.
DR   PRINTS; PR00011; EGFLAMININ.
DR   SMART; SM00181; EGF; 9.
DR   SMART; SM00180; EGF_Lam; 17.
DR   SMART; SM00281; LamB; 2.
DR   SMART; SM00282; LamG; 5.
DR   SMART; SM00136; LamNT; 1.
DR   PROSITE; PS00019; ACTININ_1; 1.
DR   PROSITE; PS00022; EGF_1; 14.
DR   PROSITE; PS01186; EGF_2; 4.
DR   PROSITE; PS01248; LAMININ_TYPE_EGF; 16.
DR   PROSITE; PS50025; LAM_G_DOMAIN; 4.
SQ   SEQUENCE   3375 AA;  374742 MW;  38CB65C01BB6E416 CRC64;

  Query Match             4.3%;  Score 9;  DB 5;  Length 3375;
  Best Local Similarity   100.0%;  Pred. No. 11;
  Matches    9;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        108 NDGEWHRVT 116
              |||||||||
Db       3089 NDGEWHRVT 3097

RESULT 4
Q81XS5
ID   Q81XS5      PRELIMINARY;      PRT;   494 AA.
AC   Q81XS5;
DT   01-JUN-2003 (TrEMBLrel. 24, Created)
DT   01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Cytosol aminopeptidase.
GN   PEPA OR BA5155.
OS   Bacillus anthracis (strain Ames).
```

```
OC   Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX   NCBI_TaxID=198094;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=22608414; PubMed=12721629;
RA   Read T.D., Peterson S.N., Tourasse N., Baillie L.W., Paulsen I.T.,
RA   Nelson K.E., Tettelin H., Fouts D.E., Eisen J.A., Gill S.R.,
RA   Holtzapple E.K., Okstad O.A., Helgason E., Rilstone J., Wu M.,
RA   Kolonay J.F., Beanan M.J., Dodson R.J., Brinkac L.M., Gwinn M.,
RA   DeBoy R.T., Madpu R., Daugherty S.C., Durkin A.S., Haft D.H.,
RA   Nelson W.C., Peterson J.D., Pop M., Khouri H.M., Radune D.,
RA   Benton J.L., Mahamoud Y., Jiang L., Hance I.R., Weidman J.F.,
RA   Berry K.J., Plaut R.D., Wolf A.M., Watkins K.L., Nierman W.C.,
RA   Hazen A., Cline R., Redmond C., Thwaite J.E., White O., Salzberg S.L.,
RA   Thomason B., Friedlander A.M., Koehler T.M., Hanna P.C., Kolsto A.-B.,
RA   Fraser C.M.;
RT   "The genome sequence of Bacillus anthracis Ames and comparison to
RT   closely related bacteria.";
RL   Nature 423:81-86(2003).
DR   EMBL; AE017040; AAP28827.1; -.
DR   TIGR; BA5155; -.
DR   GO; GO:0005622; C:intracellular; IEA.
DR   GO; GO:0004177; F:aminopeptidase activity; IEA.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR   InterPro; IPR008330; Ctsl_amnpptdse_B.
DR   InterPro; IPR000819; Peptidase_M17_C.
DR   InterPro; IPR008283; Peptidase_M17_N.
DR   Pfam; PF00883; Peptidase_M17; 1.
DR   Pfam; PF02789; Peptidase_M17_N; 1.
DR   PRINTS; PR00481; LAMNOPPTDASE.
DR   PROSITE; PS00631; CYTOSOL_AP; 1.
DR   PIRSF; PIRSF036388; Ctsl_amnpptdse_B; 1.
KW   Aminopeptidase; Complete proteome.
SQ   SEQUENCE   494 AA;  53513 MW;  D400D6CC81FA31DF CRC64;

  Query Match             3.8%;  Score 8;  DB 16;  Length 494;
  Best Local Similarity   100.0%;  Pred. No. 22;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        158 DVATLTGG 165
              ||||||||
Db        366 DVATLTGG 373

RESULT 5
Q816E3
ID   Q816E3      PRELIMINARY;      PRT;   494 AA.
AC   Q816E3;
DT   01-JUN-2003 (TrEMBLrel. 24, Created)
DT   01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Cytosol aminopeptidase (EC 3.4.11.1).
GN   BC4921.
OS   Bacillus cereus (strain ATCC 14579 / DSM 31).
OC   Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX   NCBI_TaxID=226900;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=22608415; PubMed=12721630;
RA   Ivanova N., Sorokin A., Anderson I., Galleron N., Candelon B.,
RA   Kapatral V., Bhattacharyya A., Reznik G., Mikhailova N., Lapidus A.,
RA   Chu L., Mazur M., Goltsman E., Larsen N., D'Souza M., Walunas T.,
RA   Grechkin Y., Pusch G., Haselkorn R., Fonstein M., Ehrlich S.D.,
RA   Overbeek R., Kyrpides N.;
RT   "Genome sequence of Bacillus cereus and comparative analysis with
RT   Bacillus anthracis.";
RL   Nature 423:87-91(2003).
DR   EMBL; AE017013; AAP11794.1; -.
DR   GO; GO:0005622; C:intracellular; IEA.
DR   GO; GO:0016787; F:hydrolase activity; IEA.
DR   GO; GO:0004178; F:leucyl aminopeptidase activity; IEA.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR   InterPro; IPR008330; Ctsl_amnpptdse_B.
```

```
DR   InterPro; IPR000819; Peptidase_M17_C.
DR   InterPro; IPR008283; Peptidase_M17_N.
DR   Pfam; PF00883; Peptidase_M17; 1.
DR   Pfam; PF02789; Peptidase_M17_N; 1.
DR   PRINTS; PR00481; LAMNOPPTDASE.
DR   PROSITE; PS00631; CYTOSOL_AP; 1.
DR   PIRSF; PIRSF036388; Ctsl_amnpptdse_B; 1.
KW   Aminopeptidase; Hydrolase; Complete proteome.
SQ   SEQUENCE   494 AA;  53617 MW;  09E371EC83B0FBDF CRC64;

  Query Match             3.8%;  Score 8;  DB 16;  Length 494;
  Best Local Similarity   100.0%;  Pred. No. 22;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        158 DVATLTGG 165
              ||||||||
Db        366 DVATLTGG 373


RESULT 6
Q7X2C8
ID   Q7X2C8      PRELIMINARY;      PRT;   497 AA.
AC   Q7X2C8;
DT   01-OCT-2003 (TrEMBLrel. 25, Created)
DT   01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Thermostable leucine aminopeptidase.
GN   LAP.
OS   Geobacillus kaustophilus.
OC   Bacteria; Firmicutes; Bacillales; Bacillaceae; Geobacillus.
OX   NCBI_TaxID=1462;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Lin L.-L., Hsu W.-H.;
RT   "Bacillus kaustophilus thermostable leucine aminopeptidase.";
RL   Submitted (MAY-2003) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AY308074; AAP73418.1; -.
KW   Aminopeptidase.
SQ   SEQUENCE   497 AA;  53785 MW;  2793B98222DBE695 CRC64;

  Query Match             3.8%;  Score 8;  DB 2;  Length 497;
  Best Local Similarity   100.0%;  Pred. No. 22;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        158 DVATLTGG 165
              ||||||||
Db        371 DVATLTGG 378


RESULT 7
Q8YWD5
ID   Q8YWD5      PRELIMINARY;      PRT;   497 AA.
AC   Q8YWD5;
DT   01-MAR-2002 (TrEMBLrel. 20, Created)
DT   01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT   01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE   Sodium/solute symporter.
GN   ALL1678.
OS   Anabaena sp. (strain PCC 7120).
OC   Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Nostoc.
OX   NCBI_TaxID=103690;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=21595285; PubMed=11759840;
RA   Kaneko T., Nakamura Y., Wolk C.P., Kuritz T., Sasamoto S.,
RA   Watanabe A., Iriguchi M., Ishikawa A., Kawashima K., Kimura T.,
RA   Kishida Y., Kohara M., Matsumoto M., Matsuno A., Muraki A.,
RA   Nakazaki N., Shimpo S., Sugimoto M., Takazawa M., Yamada M.,
RA   Yasuda M., Tabata S.;
RT   "Complete genomic sequence of the filamentous nitrogen-fixing
RT   cyanobacterium Anabaena sp. strain PCC 7120.";
RL   DNA Res. 8:205-213(2001).
DR   EMBL; AP003586; BAB78044.1; -.
```

```
DR   PIR; AH2015; AH2015.
DR   GO; GO:0016020; C:membrane; IEA.
DR   GO; GO:0005215; F:transporter activity; IEA.
DR   GO; GO:0006810; P:transport; IEA.
DR   InterPro; IPR001734; Na/solut_symport.
DR   Pfam; PF00474; SSF; 1.
DR   PROSITE; PS50283; NA_SOLUT_SYMP_3; 1.
KW   Complete proteome.
SQ   SEQUENCE   497 AA;  53784 MW;  DF7B1111618FE222 CRC64;

  Query Match             3.8%;  Score 8;  DB 16;  Length 497;
  Best Local Similarity   100.0%;  Pred. No. 22;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         56 STASGLLL 63
              ||||||||
Db        325 STASGLLL 332


RESULT 8
Q7TQI9
ID   Q7TQI9      PRELIMINARY;      PRT;   659 AA.
AC   Q7TQI9;
DT   01-OCT-2003 (TrEMBLrel. 25, Created)
DT   01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Hypothetical protein (Fragment).
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=FVB/N; TISSUE=Colon;
RX   MEDLINE=22388257; PubMed=12477932;
RA   Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA   Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA   Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA   Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA   Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA   Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA   Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA   Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA   Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA   Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA   Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA   Fahey J., Helton E., Ketteman M., Madan A., Rodrigues S., Sanchez A.,
RA   Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA   Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA   Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA   Krzywinski M.I., Skalska U., Smailus D.E., Schnerch A., Schein J.E.,
RA   Jones S.J., Marra M.A.;
RT   "Generation and initial analysis of more than 15,000 full-length human
RT   and mouse cDNA sequences.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=FVB/N; TISSUE=Colon;
RA   Strausberg R.;
RL   Submitted (JUN-2003) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; BC054389; AAH54389.1; -.
KW   Hypothetical protein.
FT   NON_TER       1      1
SQ   SEQUENCE   659 AA;  72436 MW;  49C86D1515E13DFB CRC64;

  Query Match             3.8%;  Score 8;  DB 11;  Length 659;
  Best Local Similarity   100.0%;  Pred. No. 29;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         48 TIELEVRT 55
              ||||||||
Db        325 TIELEVRT 332
```

```
RESULT 9
Q8R145
ID   Q8R145      PRELIMINARY;      PRT;   858 AA.
AC   Q8R145;
DT   01-JUN-2002 (TrEMBLrel. 21, Created)
DT   01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Hypothetical protein (Fragment).
GN   LAMA2.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Strausberg R.;
RL   Submitted (MAR-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; BC025503; AAH25503.1; -.
DR   MGD; MGI:99912; Lama2.
DR   GO; GO:0005604; C:basement membrane; IDA.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR001791; Laminin_G.
DR   Pfam; PF00054; laminin_G; 5.
DR   SMART; SM00282; LamG; 4.
DR   PROSITE; PS50025; LAM_G_DOMAIN; 5.
KW   Hypothetical protein.
FT   NON_TER      1      1
SQ   SEQUENCE   858 AA;  94298 MW;  0B828A266877C03A CRC64;

  Query Match             3.8%;  Score 8;  DB 11;  Length 858;
  Best Local Similarity   100.0%;  Pred. No. 36;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         48 TIELEVRT 55
              ||||||||
Db        524 TIELEVRT 531

RESULT 10
O79191
ID   O79191      PRELIMINARY;      PRT;    49 AA.
AC   O79191;
DT   01-NOV-1998 (TrEMBLrel. 08, Created)
DT   01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Nicotinamide dehydrogenase subunit 2 (Fragment).
GN   ND2.
OS   Ninox strenua (Powerful owl).
OG   Mitochondrion.
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Archosauria; Aves; Neognathae; Strigiformes; Strigidae; Ninox.
OX   NCBI_TaxID=73883;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Heart muscle;
RA   Norman J.A., Christidis L.;
RT   "Molecular data confirms the species status of the Christmas Island
RT   Hawk-Owl Ninox natalis.";
RL   Submitted (JUL-1998) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF075567; AAC34656.1; -.
DR   GO; GO:0005739; C:mitochondrion; IEA.
DR   GO; GO:0008137; F:NADH dehydrogenase (ubiquinone) activity; IEA.
DR   GO; GO:0016491; F:oxidoreductase activity; IEA.
DR   GO; GO:0006120; P:mitochondrial electron transport, NADH to u. . .; IEA.
DR   InterPro; IPR001750; Oxidored_q1.
DR   Pfam; PF00361; oxidored_q1; 1.
KW   NAD; Oxidoreductase; Ubiquinone; Mitochondrion.
FT   NON_TER      1      1
FT   NON_TER     49     49
SQ   SEQUENCE   49 AA;  5444 MW;  BF0E631BC5BF20D4 CRC64;

  Query Match             3.3%;  Score 7;  DB 8;  Length 49;
  Best Local Similarity   100.0%;  Pred. No. 31;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         57 TASGLLL 63
              |||||||
Db         20 TASGLLL 26

RESULT 11
Q7UXR0
ID   Q7UXR0      PRELIMINARY;      PRT;    71 AA.
AC   Q7UXR0;
DT   01-OCT-2003 (TrEMBLrel. 25, Created)
DT   01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Probable integrase.
GN   RB1178.
OS   Rhodopirellula baltica.
OC   Bacteria; Planctomycetes; Planctomycetacia; Planctomycetales;
OC   Planctomycetaceae; Pirellula.
OX   NCBI_TaxID=117;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=1;
RX   MEDLINE=22735913; PubMed=12835416;
RA   Gloeckner F.O., Kube M., Bauer M., Teeling H., Lombardot T.,
RA   Ludwig W., Gade D., Beck A., Borzym K., Heitmann K., Rabus R.,
RA   Schlesner H., Amann R., Reinhardt R.;
RT   "Complete genome sequence of the marine planctomycete Pirellula sp.
RT   strain 1.";
RL   Proc. Natl. Acad. Sci. U.S.A. 100:8298-8303(2003).
DR   EMBL; BX294134; CAD71943.1; -.
KW   Complete proteome.
SQ   SEQUENCE   71 AA;  7916 MW;  50D74765091B52BC CRC64;

  Query Match             3.3%;  Score 7;  DB 16;  Length 71;
  Best Local Similarity   100.0%;  Pred. No. 43;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         60 GLLLWQG 66
              |||||||
Db         12 GLLLWQG 18

RESULT 12
Q82PU0
ID   Q82PU0      PRELIMINARY;      PRT;    91 AA.
AC   Q82PU0;
DT   01-JUN-2003 (TrEMBLrel. 24, Created)
DT   01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT   01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE   Hypothetical protein.
GN   SAV782.
OS   Streptomyces avermitilis.
OC   Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC   Streptomycineae; Streptomycetaceae; Streptomyces.
OX   NCBI_TaxID=33903;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=MA-4680 / ATCC 31267 / NCIMB 12804 / NRRL 8165;
RX   MEDLINE=21477403; PubMed=11572948;
RA   Omura S., Ikeda H., Ishikawa J., Hanamoto A., Takahashi C.,
RA   Shinose M., Takahashi Y., Horikawa H., Nakazawa H., Osonoe T.,
RA   Kikuchi H., Shiba T., Sakaki Y., Hattori M.;
RT   "Genome sequence of an industrial microorganism Streptomyces
RT   avermitilis: deducing the ability of producing secondary
RT   metabolites.";
RL   Proc. Natl. Acad. Sci. U.S.A. 98:12215-12220(2001).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=MA-4680 / ATCC 31267 / NCIMB 12804 / NRRL 8165;
RX   MEDLINE=22608306; PubMed=12692562;
RA   Ikeda H., Ishikawa J., Hanamoto A., Shinose M., Kikuchi H., Shiba T.,
```

```
RA   Sakaki Y., Hattori M., Omura S.;
RT   "Complete genome sequence and comparative analysis of the industrial
RT   microorganism Streptomyces avermitilis.";
RL   Nat. Biotechnol. 21:526-531(2003).
DR   EMBL; AP005024; BAC68492.1; -.
KW   Hypothetical protein; Complete proteome.
SQ   SEQUENCE   91 AA;  9926 MW;  73D27F5589F8FEAD CRC64;

  Query Match             3.3%;  Score 7;  DB 16;  Length 91;
  Best Local Similarity   100.0%;  Pred. No. 54;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        119 REGRRGS 125
              |||||||
Db         67 REGRRGS 73


RESULT 13
P74056
ID   P74056      PRELIMINARY;      PRT;    93 AA.
AC   P74056;
DT   01-FEB-1997 (TrEMBLrel. 02, Created)
DT   01-FEB-1997 (TrEMBLrel. 02, Last sequence update)
DT   01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE   Hypothetical protein ssr1391.
GN   SSR1391.
OS   Synechocystis sp. (strain PCC 6803).
OC   Bacteria; Cyanobacteria; Chroococcales; Synechocystis.
OX   NCBI_TaxID=1148;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=97061201; PubMed=8905231;
RA   Kaneko T., Sato S., Kotani H., Tanaka A., Asamizu E., Nakamura Y.,
RA   Miyajima N., Hirosawa M., Sugiura M., Sasamoto S., Kimura T.,
RA   Hosouchi T., Matsuno A., Muraki A., Nakazaki N., Naruo K., Okumura S.,
RA   Shimpo S., Takeuchi C., Wada T., Watanabe A., Yamada M., Yasuda M.,
RA   Tabata S.;
RT   "Sequence analysis of the genome of the unicellular cyanobacterium
RT   Synechocystis sp. strain PCC6803. II. Sequence determination of the
RT   entire genome and assignment of potential protein-coding regions.";
RL   DNA Res. 3:109-136(1996).
DR   EMBL; D90911; BAA18132.1; -.
DR   PIR; S75571; S75571.
KW   Hypothetical protein; Complete proteome.
SQ   SEQUENCE   93 AA;  10746 MW;  7206B38D3AF74096 CRC64;

  Query Match             3.3%;  Score 7;  DB 16;  Length 93;
  Best Local Similarity   100.0%;  Pred. No. 55;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         48 TIELEVR 54
              |||||||
Db         35 TIELEVR 41


RESULT 14
Q9XHS4
ID   Q9XHS4      PRELIMINARY;      PRT;   106 AA.
AC   Q9XHS4;
DT   01-NOV-1999 (TrEMBLrel. 12, Created)
DT   01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Hypothetical protein (Fragment).
GN   FAD81.
OS   Brassica campestris (Field mustard).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC   eurosids II; Brassicales; Brassicaceae; Brassica.
OX   NCBI_TaxID=3711;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=cv. R500;
RA   Brunel D., Froger N., Pelletier G.;
RT   "Development of amplified consensus genetic markers (A.C.G.M.) in
RT   Brassica napus from Arabidopsis thaliana sequences of known biological
RT   function.";
RL   Submitted (APR-1998) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF056568; AAD41579.1; -.
DR   GO; GO:0016491; F:oxidoreductase activity; IEA.
DR   InterPro; IPR005804; FA_desat_fam.
DR   Pfam; PF00487; FA_desaturase; 1.
DR   ProDom; PD001081; FA_desat_fam; 1.
KW   Hypothetical protein.
FT   NON_TER       1      1
FT   NON_TER     106    106
SQ   SEQUENCE   106 AA;  11821 MW;  B05907C7957E6903 CRC64;

  Query Match             3.3%;  Score 7;  DB 10;  Length 106;
  Best Local Similarity   100.0%;  Pred. No. 62;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        184 RPGAPPP 190
              |||||||
Db          1 RPGAPPP 7


RESULT 15
Q9XHS6
ID   Q9XHS6      PRELIMINARY;      PRT;   108 AA.
AC   Q9XHS6;
DT   01-NOV-1999 (TrEMBLrel. 12, Created)
DT   01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Hypothetical protein (Fragment).
GN   FAD81.
OS   Brassica oleracea (Cauliflower).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC   eurosids II; Brassicales; Brassicaceae; Brassica.
OX   NCBI_TaxID=3712;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=cv. Rapide Cycling;
RA   Brunel D., Froger N., Pelletier G.;
RT   "Development of amplified consensus genetic markers (A.C.G.M.) in
RT   Brassica napus from Arabidopsis thaliana sequences of known biological
RT   function.";
RL   Submitted (APR-1998) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF056566; AAD41577.1; -.
DR   GO; GO:0016491; F:oxidoreductase activity; IEA.
DR   InterPro; IPR005804; FA_desat_fam.
DR   Pfam; PF00487; FA_desaturase; 1.
DR   ProDom; PD001081; FA_desat_fam; 1.
KW   Hypothetical protein.
FT   NON_TER       1      1
FT   NON_TER     108    108
SQ   SEQUENCE   108 AA;  12075 MW;  282D56101E045E83 CRC64;

  Query Match             3.3%;  Score 7;  DB 10;  Length 108;
  Best Local Similarity   100.0%;  Pred. No. 63;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        184 RPGAPPP 190
              |||||||
Db          3 RPGAPPP 9


Search completed: March  9, 2004, 17:30:35
Job time : 19.3903 secs
```